REiD (RESEARCH AND EVALUATION IN EDUCATION)
Vol. 5, No. 1, June 2018

Indexed in:

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS | ISJD | Google Scholar | sinta Science and Technology Index

9 772460 699001

Research and Evaluation in Education

Vol. 5, No. 1, June 2019

Research and Evaluation in Education

Vol. 5, No. 1, June 2019

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

Editor in Chief
Djemari Mardapi, *Universitas Negeri Yogyakarta*

Editors
Badrun Kartowagiran, *Faculty of Engineering, Universitas Negeri Yogyakarta*
Edi Istiyono, *Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta*
Samsul Hadi, *Faculty of Engineering, Universitas Negeri Yogyakarta*
Elizabeth Hartnell-Young, *Melbourne Graduate School of Education, University of Melbourne, Australia*
John Hope, *Faculty of Education, University of Auckland, New Zealand*
Suzanne Rice, *Graduate School of Education, University of Melbourne, Australia*
Nur Hidayanto Pancoro Setyo Putro, *Faculty of Languages and Arts, Universitas Negeri Yogyakarta*
Risky Setiawan, *Faculty of Social Sciences, Universitas Negeri Yogyakarta*
Syukrul Hamdi, *Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta*
Alita Arifiana Anisa, *Lembaga Pendidikan dan Pengembangan Profesi Indonesia*
Lye Che Yee, *Singapore Examinations and Assessment Board, Singapore*
Socheath Mam, *Faculty of Education, Royal University of Phnom Penh, Cambodia*

Journal Coordinator of Graduate School of Universitas Negeri Yogyakarta
Heri Retnawati

Setting
Ririn Susetyaningsih
Syarief Fajaruddin
Syahri Ramadan

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
2nd Floor of General Lecture Building, Jl. Colombo No. 1, Karangmalang, Depok, Sleman, Yogyakarta 55281
Telephone: 0274 586168 ext. 229 or 0274 550836, Facsimile: 0274 520326
E-mail: reid.ppsuny@uny.ac.id, reid.ppsuny@gmail.com

Copyright © 2019, REiD (Research and Evaluation in Education)

Foreword

We are very pleased that REiD (Research and Evaluation in Education) is releasing its ninth edition. We are also very excited that the journal has been attracting papers from the neighbouring country, Cambodia. 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 common education, but also comprises the results of research in education in a broader coverage, such as natural sciences, mathematics, language education, social sciences, and communication program, 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 issue.

Yogyakarta, June 2019

Editor in Chief

TABLE OF CONTENT

REiD (Research and Evaluation in Education), 5(1), 2019, 1-9
Available online at: http://journal.uny.ac.id/index.php/reid

Developing psychomotor evaluation instrument of biochemistry practicum for university students of biology education

[*1]Etika Dyah Puspitasari; [2]Mohamad Joko Susilo; [3]Novi Febrianti
[1,2,3]Department of Biology Education, Universitas Ahmad Dahlan
Jl. Prof. Dr. Soepomo, S.H., Janturan, Yogyakarta 55164, Indonesia
[*]Corresponding Author. E-mail: etika.puspitasari@pbio.uad.ac.id

Submitted: 04 December 2018 | Revised: 06 December 2018 | Accepted: 06 December 2018

Abstract

Practicum is one of the important aspects of the learning of biology. There is no psychomotor evaluation instrument that is valid and reliable. This study is aimed at developing a valid and reliable psychomotor evaluation instrument for biochemistry practicum. The study is developmental research using the 4-D model of 'define, design, develop, and disseminate'. Instrument validation was carried out through construct validation. The findings show that the developed instrument is characterized by a high level of construct validity although the reliability measure is not very well-estimated. The instrument is constructed of four factors of perception, set, guided response, and mechanism developed into 80 statement items.

Keywords: *instrument development, psychomotor domain, practicum, biochemistry, biology education*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.22126

Introduction

Science learning (particularly biology) involves practicum. A practicum, in biology learning, is an activity of exploration as well as experimentation in the laboratory or in the open field to give a direct experience to the students. Since science covers three aspects: product, process, and scientific attitudes (Tursinawati, 2016), it is often stated that practicum is an unseparated part of science.

One of the superior aspects of practicum as a learning method is that it gives the chance for students to test, find, and elucidate theories (Suryaningsih, 2017); develop the basic skills of experimentation; endorse enthusiasm for knowledge; elevate problem-solving skills; provide students with facilities of scientific investigation. Practicum activities can also improve the students' scientific processes and concept masteries (Lestari & Diana, 2018; Suardana, Liliasari, & Ismunandar, 2013).

It is unfortunate to say that, thus far, evaluation on practicum activities in the laboratory still emphasize on the cognitive aspects, while psychomotor skills evaluation receives small attention (Hamid et al., 2012). This can be seen from the low proportion of the cognitive evaluation for the pre-test, post-test, and final assignment that is usually written. Meanwhile, according to Osman, Hiong, and Vebrianto (2013), in order that students acquire the skills needed for the 21st-century, biology learning must involve a lot of inquiry skills. Inquiry skills include (1) formulating problems, (2) proposing a hypothesis, (3) designing experimentation to test hypotheses, (4) testing data analyses and making conclusions, and (5) writing a report. In addition, students are also expected to be able to operate experiment tools in the laboratory.

Maknun, Surtikanti, and Subahar (2012) map the essential laboratory skills into 14 as follows: (1) observing, (2) calculating, (3) measuring, (4) classifying, (5) finding space/time relation, (6) formulating hypotheses, (7) designing an experiment, (8) controlling variables, (9) interpreting data, (10) making infer-

This is an open access article under the CC-BY-SA license.

ences, (11) predicting, (12) concluding, (13) applying, and (14) communicating. However, in their study, they find that mastery of the essential laboratory skills of biology teacher-candidate students in the ecology practicum is still low.

Likewise, Kasilingam, Ramalingam, and Chinnavan (2014) describe psychomotor skills into seven levels, namely: (1) perception, (2) set, (3) guided response, (4) mechanism, (5) complex overt, (6) adaptation, and (7) origination. Verbs that are related to the perception level include selecting, choosing, isolating, and identifying. Verbs that represent the set level include showing, starting, explaining, etc. In the guided response level, verbs that can be used include imitating, following steps, making, etc. In the mechanism level, verbs that are relevant include calibrating measuring, mixing, organizing, heating, manipulating, etc.

Maknun, Surtikanti, Munandar, and Subahar (2012) categorize psychomotor skills of the practicum class in the ecology subject matter as setting up the tools in line with the practicum plan, calibrating and maintaining the laboratory tools, operating pipettes, operating microscopes, taking notes, working safely in accordance with work health and security. The results of their study show that the psychomotor skills of teacher-candidate students in biology practicum are still low.

The conduct of the biochemistry practicum in the study program of Biology Education, Universitas Ahmad Dahlan (UAD) has included cognitive, psychomotor, and affective aspects; however, evaluation of the cognitive aspects dominates the process (80%) while the psychomotor and affective aspects take the rest (20%). Besides, no standard and valid instrument have been developed for the evaluation of the psychomotor aspects of learning in the biochemistry practicum. As a result, evaluation for the practicum has a high measure of subjectivity.

To date, many studies have been conducted on the development of evaluation instruments. One example is the study done by Ridlo (2012), but this study focuses on the knowledge aspects of biology practicum. Instrument development in the psychomotor domain in biology practicum is done, among others, by Yunita, Agung, and Nuraeni (2016) with good validation in the aspects of material, construction, and language. Yulianti, Andriani, and Taufiq (2014) have developed a psychomotor evaluation instrument in the temperature and calorie topic. Another study was done by Hazarianti, Masriani, and Hadi (2016) on a psychomotor evaluation rubric in the practicum of the distribution coefficient sub-material. This rubric, however, is used in the classes other than biology.

Development of psychomotor evaluation instruments has been done so far for high school students; meanwhile, very little has been done for university students. Besides, learning evaluation has so far emphasized the cognitive skills, even for practicum classes which actually need psychomotor skills. It is therefore important that the development of evaluation instruments be developed in biology education, especially in the biochemistry topic. This is due to the fact that biochemistry is one of the basic materials in biology along with physiology, genetics, microbiology, and others.

Method

This study is a research and development using the four-D model proposed by Thiagarajan, Semmel, and Semmel (1974). The four developmental phases in the model are as follows: (a) define, (b) design, (c) develop, and (d) disseminate (Haviz, 2013). This study is con-fined up to the 'develop' phase, leaving out the 'disseminate' phase (Febriana, Rachmadiarti, & Faizah, 2016; Noverina, Taufiq, & Wiyono, 2014; Yunita et al., 2016).

The 'define' phase includes four steps, namely: (a) initial analysis, (b) curriculum review, (c) content review, and (d) learner analysis. The 'design' phase includes four steps, namely: (a) selection of assessment scales, (b) development of the instrument draft, (c) instrument validation, and (d) test of assistant limitation. The 'develop' phase consists of two steps, namely: (a) product evaluation by experts and (b) small-group and large-group tryouts.

The study used two questionnaires as the instruments for data collection. The first questionnaire, using the Likert scale, consisted

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

of statements concerning the instrument feasibility to be given to material experts and evaluation experts. The second questionnaire was tested for readability to practicum assistants using the Guttman (Yes/No) scale. All instruments were first validated by the material and evaluation experts.

Data were analyzed by a combination of descriptive and qualitative techniques. The Likert scale was scored by 4 to 1 rating to be categorized into very good, good, poor, and very poor. The Guttman scale had 2 ratings with a maximum score of 15. For item validity, exploratory factor analysis (EFA) was used with the four indicators of Kaiser-Meyer-Olkin measure of sampling adequacy (KMO, MSA), Bartlett's test of sphericity, anti-image correlation, and factor loading. As general criteria, if the level of Bartlett's test of sphericity is $p<0.5$, KMO-MSA value is >0.5, and the anti-image correlation is >0.5, the sample data are feasible for analysis. The quantitative data from the experts and assistants were analyzed for feasibility by categorizing them into four interpretation criteria using the formula proposed by Mardapi (2008).

The research product is regarded as feasible if the results of the analyses are minimally at the category of 'good'. The criteria include content material, construction, language, objectivity, and utility.

Findings and Discussion

Findings

The study is research and development in three phases, namely: (a) define, (b) design, and (c) develop. In the 'define' phase, analyses are conducted in the initial situation, curriculum content, subject material, and learner characteristics. Analyses of the initial situation are done by carrying out discussions with biochemistry practicum coordinators and assistants concerning the running and evaluation of the biochemistry practicum. From this activity, it can be known that the practice of practicum evaluation is still dominant in the cognitive domain, approaching 80% of the whole process. Psychomotor skills aspects take only about 10%.

The curriculum analyses are done on the practicum lesson plans, learning outcomes, and practicum guidebooks. Concerning the learning outcomes, among others, students are able to practice making pH solutions of various concentrations, making buffer solutions, and measuring pH solutions. These abilities in making and measuring pH solutions will become the bases for doing other practicum activities.

Analyses of the content material are directed to look at the basic materials that are given before the practicum class. The content material for pH practicum is an advanced topic. The topic of pH making and measuring are the fifth items in the whole syllabus of the biochemistry practicum. The preceding classes contain practicum activities the accuracy and correctness of experiments. In these preceding practicums, students practice liquefying, measuring, and using the right tools. It is expected that in the fifth practicum, students are readily familiar with the initial and basic steps of experimentation.

Learner analyses are directed to look at the characteristics of the students who take the biochemistry practicum in semester 2. The biochemistry practicum is the first practicum the students have in their program. There is no practicum in semester 1. The practicum uses four of the six levels of the psychomotor domain (Hamid et al., 2012) namely: level 1 (perceiving), level 2 (being ready for active participation), level 3 (integrative responding), and level 4 (showing work performances to become habitual). In the complete scheme, level 5 is complex overt responding and level 6 is adapting. These are not yet included in the items for the learning evaluation.

In the 'design' phase, the following steps were carried out: selecting evaluation scales, developing the instrument draft, validating, and readability testing. Selection of the evaluation scales is done by reviewing the instrument draft design. Initially, the evaluation scales are related to the check-list type with Yes/No responses. According to Ibezim and Igwe (2016), the check-list instrument is more objective in measuring psychomotor skills than rating scales. However, taking the experts and assistants' recommendation, the

Likert-type rating scales be used for the developed instrument. It is expected that, by using the Likert scales, differences in the students' performances can be more clearly detected. Three scales will be used: 1 for inadequate, 2 for good, and 3 for very good.

The product draft consists of an instrument for readability and instrument for expert validation. The draft is formatted in the following aspects: (1) title of the experiment, (2) objectives to be achieved, (3) psychomotor evaluation aspects, (4) levels of the psychomotor skills, (5) indicators for the psychomotor skills, (6) descriptors representing the indicators, (7) evaluation scales, (8) evaluation rubrics, and (9) scoring guides.

The experiment title is related to the experiment of making and measuring pH. The learning objective to be achieved is for students to be able to make solutions with various concentrations, making solution buffers, and measuring pH solutions. The aspects that will be observed in the activities consist of preparation for the practicum, running of the practicum, and reporting.

The product instrument was evaluated by validators before it was subjected to the try-outs. This evaluation consists of readability checks by practicum assistants and evaluation instrument by evaluation and subject matter experts.

The instrument validity was obtained from the wider-scale try-out using exploratory factor analysis (EFA). The results of the EFA analyses show that the Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO) is 0.787 which means that the factor analysis can be continued. Looking at the number of factors that have an eigenvalue of more than 1, four levels of the psychomotor domain can be obtained: level 1 for perception, level 2 for set (readiness for active participation), level 3 for guided response (integrative responses), and level 4 for mechanism (showing performance as a habit).

The indicators for the psychomotor skills in the developed instrument cover the following details: being able to set the tools and materials for the experiment, writing up the steps of the work, making HCl solution using various concentrations, measuring the pH of the HCl solution, making 2% NaOH solution, making 100 ml of 0.2M CH3COONa solution, making 1% gelatin solution, making 0.2M NaH2PO4.H2O solution, making 0.2M pH 5 acetate buffer solution, writing out practicum objectives, writing out observation results, comparing observation results with the theory, writing out the discussion of results, making conclusions, and writing up the practicum report. These are presented in Table 1. Each indicator is operationalized into descriptors. There are 80 descriptor statements. The three three-scale Likert criteria are 1 for inadequate, 2 for good, and 3 for very good.

Table 1. Psychomotor levels and indicators of the instrument

Psychomotor Level	Indicator
Perception	Selecting tools and materials
	Formulating practicum objectives
Set	Writing up sequence of works
	Writing out results of observation
Guided responses	Making *HCl* solution using various concentrations
	Making 2% *NaOH* solution
	Making 100 ml of 0.2M *CH3COONa* solution
	Making 1% gelatin solution
	Making 0.2M *NaH2PO4.H2O* solution
	Making 0.2M pH 5 acetate buffer solution
	Comparing observation results with the theory
	Writing out the discussion of results
	Making conclusions
	Writing up the practicum report
Mechanism	Calculating the solution volume
	Weighing materials
	Determining height or volume of solution using practicum tools (pippete, bulb, glass, etc.)
	Measuring *pH* solution

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

The total score made by the students is the sum of all the scores obtained for each indicator. The maximum score is 240 and the minimum score is 80. Students' score can be obtained by the following formula:

$$\text{Student's score} = \frac{\text{Score gained by student}}{240} \times 100$$

The instrument that has been construct-validated was subjected to readability checks by the practicum assistants. The results of the readability test show that the instrument readability can be categorized as very good (93.83%). The readability checks include language, ease, objectivity, and utility.

The 'develop' phase consists of three activities, covering: (a) expert evaluation, (b) small-group try-out, and also (c) large-group try-out. Based on the results of the evaluation by the subject-matter and evaluation experts, the instrument is categorized as very good (91.67). The evaluation includes language, construct, content, objectivity, and utility. A minor revision is suggested, however, by the subject-matter experts on the use of vocabulary words and simplification of the descriptors. The final version of the instrument draft ends up with 80 statement items. Some indicators and descriptors of the final draft are presented in Table 2, Table 3, Table 4, and Table 5.

In Table 2, the indicators and descriptors are those that are used for the perception level. There are two indicators in this level, namely selecting tools and materials and formulating practicum objectives. These indicators are supported by Kasilingam et al. (2014) whereby the level perception can be operationalized by choosing, selecting, describing, etc.

Table 3 shows indicators and descriptors for the psychomotor set level. The set level operationalizes into mental, emotional, and physical readiness of the student to work. In this level, the indicators are writing down work procedure and writing up observation results. These indicators are chosen for the reason that students' readiness to do the practicum can be seen from their understanding on the sequence of the steps in the practicum class, which is represented by their ability to write down the steps in accordance with the guidebook. In the same way, students' readiness to communicate the results and write a report is shown by their ability to write down the results of the practicum and any important phenomenon in the form of a report draft.

Table 2. Indicators and descriptors operationalized from perception

Indicator	Descriptor
Selecting tools and materials	Selecting tools and materials for making solution from various concentrations, making buffer solution, and measuring *pH* Arranging tools and materials on the operation desk thoroughly as directed by the guidebook
Formulating practicum objectives	Writing up learning objectives in accordance with the learning outcomes in biochemistry practicum

Table 3. Indicators and descriptors operationalized from the set level

Indicator	Descriptor
Writing down the sequence of work steps	Writing down the complete steps of the practicum job in accordance with the guidebook Arranging the tools and materials on the work desk in accordance with the guidebook
Writing up observation results	Writing up results of the practicum in a tentative draft

In the level of guided response, there are 50 items to be tested. These items are written out in accordance with the practicum guidebook (Kasilingam et al., 2014). Some of these instrument items are shown in Table 4.

The mechanism level has skill categories with which students are familiar. It includes calculating solution volume, weighing materials, observing solution volume through the glass tube, heating solution, measuring pH solution, etc. These descriptors use operational mechanisms like measuring, organizing, heating, etc. (Kasilingam et al., 2014).

After being evaluated by experts, the instrument was subjected to a try-out to a small group of 20 students. The results show that the average of students' scores is 128. Converted into the 1 to 100, this score is represented by 53.33. This score belongs to the low category.

Table 4. Indicators and descriptors operationalized from guided response

Indicator	Descriptor
Making *HCl* solution using various concentrations	Putting solution *HCl* 1M using correct pippete into 25 ml bulb tube to make *HCl* 0.05M Pouring aquades to liquefy *HCl* 1M through the bulb glass wall at correct limit stripe
Making 100 ml of 0.2M *CH3COOna* solution	Putting 1.64 gram CH3COONa into bulb tube 50ml using glass bicker Pouring aquades into bulb glass containing 1.64 gram *CH3COONa* through tube wall Closing bulb glass and shaking it *CH3 COONa* crystal solves
Making 100ml of 1% gelatin solution	Putting 1 gram gelatin into bicker glass Pouring 60 ml aquades in glass bicker containing 1 gram gelatin Solving gelatin solution Cooling gelatin solution to room temperature Pouring cool gelatin solution into 100ml bulb tube using glass bicker Pouring aquades into bulb tube to limit stripe Closing bulb tube and shaking it to make solution homogeneous
Making 100ml of 0.2M pH 5 acetate buffer solution	Taking 62.95 ml of *CH3COONa* solution Taking 37.05 ml of 0.2M *CH3COOH* solution Mixing *CH3COONa* and CH3COOH solutions by shaking Erlenmeyer to make solution homogeneous

Table 5. Indicators and descriptors operationalized from mechanism level

Indicator	Descriptor
Calculating solution volume	Constructing formulas for calculating volumes solution concentrations to be used Calculating solution volume for solution liquidation following the volume comparison formula and practicum concentration
Weighing material	Putting filtering paper or watch Glasson the analytic plate or pan Pushing marking button to calibrate scales till zero number (0) appears Putting material on filtering paper or watch glass Weighing weight of material in accordance with needed weight
Determining height or volume of solution using practicum tools (pippete, bulb, glass, etc.)	Observing height of solution volume to be measured parallel with eyes Using meniscus point to determine volume of solution Taking 5 ml of buffer solution into test tube using pipette for each solution concentration
Measuring solution *pH*	Merge tip of *pH* indicator into test tube in 5 seconds Compare colour of merged pH indicator paper with that of universal standard *pH* indicator

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

The instrument was finally subjected to the bigger-group try-out of 45 students. The results of the try-out show that the average of students' scores is 148.8. Converted into the 1 to 100, this score is represented by 62. This score also belongs to the low category.

Discussions

The research findings show that the developed instrument has a high construct validity; however, results of the small-group and large-group try-outs are not satisfactory. This may be due to the condition that the results of the practicum experiment are shared by the students in the group so that not every student is able to carry out all of the assessment aspects in the practicum.

The results of the try-out to the large group show a score that is interpretable into the low category; the same with those of the small group try-out. This may be caused by the fact that the practicum is carried out by task assignments. This was done because the practicum material is big in volume while the time is limited to two hours. This causes the condition that students are not able to conduct all the activities in the practicum so that the observed psychomotor scores are partial.

The low level of the results of the try-outs may also be caused by the suspicion that the instrument reliability measure is not very well-defined or estimated. According to Lee, Brennan, and Kolen (2000), when the reliability measure is low, the standard error of measurement (SEM) is also low; bringing about the consequence that the validity of the measurement is zero. On the other hand, when the reliability measure is high and the SEM is low, it means that there is validity in the results of the measurement. In spite of all that, the height of the reliability measures (regardless of the sizes) does not guarantee the presence of validity (Azwar, 2008). Consequently, it is true that the conduct of reliability estimation is important in instrument development.

Conclusion and Suggestions

Conclusion

Based on the research findings, it can be concluded that the developed instrument is feasible to be used. The instrument has a high measure of construct validity although its reliability is not very well-estimated. In fact, instrument reliability can be elevated in two ways, i.e. by increasing items that have high internal consistency or reducing those with low internal consistency. The instrument is constructed of four psychomotor aspects of perception, set, guided responses, and mechanism distributed into 80 statement items.

Suggestions

The developed psychomotor evaluation instrument has not been estimated very well in terms of its reliability. It is suggested that other studies intended to develop an evaluation instrument carry out reliability estimation. The techniques can be suited to the objectives and types of data of the study.

Acknowledgment

The authors thank LPPM UAD for providing the internal fund through *Hibah Penelitian Dosen Pemula* in 2018, and they deliver their gratitude to all academicians of Biology Education Study Program at Universitas Ahmad Dahlan who gave a hand in the research.

References

Azwar, S. (2008). *Penyusunan skala psikologi*. Yogyakarta: Pustaka Pelajar.

Febriana, M., Rachmadiarti, F., & Faizah, U. (2016). Kelayakan perangkat penilaian materi ekologi yang sesuai dengan tagihan kurikulum 2013. *BioEdu: Berkala Ilmiah Pendidikan Biologi*, *5*(1), 49–54.

Hamid, R., Baharom, S., Hamzah, N., Badaruzzaman, W. H. W., Rahmat, R. A. O. K., & RaihanTaha, M. (2012). Assessment of psychomotor domain in materials technology laboratory work. *Procedia - Social and Behavioral Sciences*, *56*, 718–723. https://doi.org/10.1016/J.SBSPRO.2012.09.708

Haviz, M. (2013). Research and development: Penelitian di bidang kependidikan yang inovatif, produktif dan bermakna. *Ta'dib*, *16*(1), 28–43.

Hazarianti, P., Masriani, & Hadi, L. (2016). Pengembangan rubrik penilaian psikomotorik pada praktikum submateri koefisien distribusi mahasiswa pendidikan kimia. *Jurnal Pendidikan Dan Pembelajaran*, *5*(11), 1–10.

Ibezim, N. E., & Igwe, N. (2016). Checklist versus rating scale in psychomotor assessment: Achieving objectivity. *International Journal of Humanities and Social Science Invention*, *5*(8), 48–51. https://doi.org/10.3389/fpsyt.2014.00196

Kasilingam, G., Ramalingam, M., & Chinnavan, E. (2014). Assessment of learning domains to improve student's learning in higher education. *Journal of Young Pharmacists*, *6*(1), 27–33. https://doi.org/10.5530/jyp.2014.1.5

Lee, W., Brennan, R. L., & Kolen, M. J. (2000). Estimators of conditional scale-score standard errors of measurement: A simulation study. *Journal of Educational Measurement*, *37*(1), 1–20.

Lestari, M. Y., & Diana, N. (2018). Keterampilan proses sains (KPS) pada pelaksanaan praktikum Fisika Dasar I. *Indonesian Journal of Science and Mathematics Education*, *1*(1), 49–54. Retrieved from http://ejournal.radenintan.ac.id/index.php/IJSME/article/view/2474/1828

Maknun, D., Surtikanti, R. R. H. K., Munandar, A., & Subahar, S. (2012). Keterampilan esensial dan kompetensi motorik laboratorium mahasiswa calon guru biologi dalam kegiatan praktikum ekologi. *Jurnal Pendidikan IPA Indonesia*, *1*(2), 141–148. https://doi.org/10.15294/jpii.v1i2.2131

Maknun, D., Surtikanti, R. R. H. K., & Subahar, T. S. (2012). Pemetaan keterampilan esensial laboratorium dalam kegiatan praktikum ekologi. *Jurnal Pendidikan IPA Indonesia*, *1*(1), 1–7.

Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.

Noverina, S., Taufiq, & Wiyono, K. (2014). Pengembangan rubrik penilaian keterampilan dan sikap ilmiah mata pelajaran fisika kurikulum 2013 di kelas X sekolah menengah atas. *Jurnal Inovasi Dan Pembelajaran Fisika*, *1*(2), 145–151. Retrieved from https://ejournal.unsri.ac.id/index.php/jipf/article/view/1804/749

Osman, K., Hiong, L. C., & Vebrianto, R. (2013). 21st century biology: An interdisciplinary approach of Biology, Technology, Engineering and Mathematics Education. *Procedia - Social and Behavioral Sciences*, *102*, 188–194. https://doi.org/10.1016/J.SBSPRO.2013.10.732

Ridlo, S. (2012). Pengembangan tes pengetahuan praktikum biologi berdasarkan Graded Response dan Generalized Partial Credit. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *16*(Edisi Dies Natalis ke-48 UNY), 166–182. https://doi.org/10.21831/pep.v16i0.1111

Suardana, I. N., Liliasari, L., & Ismunandar, I. (2013). Peningkatan penguasaan konsep mahasiswa melalui praktikum elektrolisis berbasis budaya lokal. *Jurnal Pendidikan Dan Pembelajaran (JPP)*, *20*(1), 45–52. Retrieved from http://journal.um.ac.id/index.php/pendidikan-dan-pembelajaran/article/view/3869

Suryaningsih, Y. (2017). Pembelajaran berbasis praktikum sebagai sarana siswa untuk berlatih menerapkan keterampilan proses sains dalam materi biologi. *Jurnal Bio Education*, *2*(2), 49–57. Retrieved from https://jurnal.unma.ac.id/index.php/BE/article/view/759/708

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Bloomington, IN: Indiana University.

Tursinawati. (2016). Penguasaan konsep hakikat sains dalam pelaksanaan percobaan pada pembelajaran IPA di SDN Kota Banda Aceh. *Jurnal Pesona Dasar*, *2*(4), 72–84. Retrieved from

http://www.jurnal.unsyiah.ac.id/PEAR/article/view/7534

Yulianti, N., Andriani, N., & Taufiq. (2014). Pengembangan instrumen penilaian psikomotorik pada pokok bahasan suhu dan kalor di SMP. *Jurnal Inovasi Dan Pembelajaran Fisika*, *1*(2), 152–156. Retrieved from http://ejournal2.unsri.ac.id/index.php/jipf/article/view/1805/750

Yunita, L., Agung, S., & Nuraeni, R. (2016). Pengembangan instrumen penilaian aspek psikomotorik siswa SMA/MA pada praktikum titrasi asam basa. In *Pros. Semnas Pend. IPA Pascasarjana UM* (pp. 662–670). Malang: Universitas Negeri Malang. Retrieved from http://pasca.um.ac.id/wp-content/uploads/2017/02/Luki-Yunita-662-670.pdf

REiD (Research and Evaluation in Education), 5(1), 2019, 10-20
Available online at: http://journal.uny.ac.id/index.php/reid

Methods used by mathematics teachers in developing parallel multiple-choice test items in school

[*1]Kartika Pramudita; [2]R. Rosnawati; [3]Socheath Mam
[1,2]Department of Educational Research and Evaluation,
Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Karangmalang, Depok, Sleman, Yogyakarta 55281, Indonesia
[3]Faculty of Education, Royal University of Phnom Penh, Cambodia
Russian Federation Boulevard, Toul Kork, Phnom Penh, Cambodia
[*]Corresponding Author. E-mail: kartika_pramudita.2017@student.uny.ac.id

Submitted: 04 December 2018 | Revised: 12 February 2019 | Accepted: 14 February 2019

Abstract

The study was aimed at describing five methods of the development of parallel test items of the multiple-choice type in mathematics at Yogyakarta (primary education level). The study was descriptive research involving 22 mathematics teachers as the respondents. Data collection was conducted through interviews and document reviews concerning the developed test packages. A questionnaire was used to gather data about the procedure the teachers employed in developing the tests. Findings show that the teachers used five methods in developing the test item; namely (1) randomizing the item numbers; (2) randomizing the sequences of response options; (3) writing items using the same contexts but different figures; (4) using anchor items; and (5) writing different items based on the same specification table. All of the respondents stated that they developed the table of the specification before developing the test items and that most of them (77%) did the validation of the instruments in content and language.

Keywords: *parallel test items, test item development, mathematics evaluation, multiple-choice testing*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.22219

Introduction

Evaluation is one of the essential aspects of education that will contribute to the achievement of educational quality. One of the objectives of evaluation is to know students' real competence. Effective evaluation can differentiate between high- and low-achieving students. An effective evaluation gathers evidences that are valid concerning learning outcome. The process and product of evaluation are also able to give improvement to students' motivation and achievement in learning (Stiggins & Chappuis, 2012, p. 3). One type of evaluation conducted in school is cognitive evaluation. Cognitive evaluation can be performed by using tests that will show the individual or group characteristics (Rasyid & Mansur, 2008, p. 11).

Assessment for learning is integral to best practice in teaching and learning. The development of a measurable test instrument must be done through qualitative and empirical research. According to Mardapi (2008, p. 15), a test instrument, either test or non-test, must have evidence for validity and reliability so that test results can be comparable and economical. A test is said to be valid if it measures what it is supposed to measure. A test with high validity will have a low error of measurement, meaning that the scores obtained by testees are close to the original scores. A test is said to be reliable if the observed scores have a high correlation with the original scores. Sources for an instrument validity can be traced from the contents of the test, in the forms of qualitative analyses of the materials, constructs, and language of the test.

This is an open access article under the CC-BY-SA license.

A test battery used in an evaluation can be in various test items. The test form selected must be in line with the objective of the testing. One common test form is the multiple-choice test. A multiple choice test item consists of a stem followed by several alternative responses (Kehoe, 1995b, p. 2). The multiple-choice test form is suitable for testing that involves an enormous amount of material, such as the national examination (NE) or national-standard school examination (NSSE). This is the superiority of multiple-choice testing in that it covers a high number of items, is objective, is efficient, and can be highly reliable (Reynolds, Livingston, & Willson, 2009, pp. 184–186). A multiple-choice test can measure all the thinking processes in the cognitive domain from the lowest to the highest levels. This can be highly suitable for testing in the field of mathematics (Torres, Lopes, Babo, & Azevedo, 2011, p. 11). A number of studies have been done for the evaluation of mathematics learning using the multiple-choice test mode. One study is conducted to measure the high-order thinking skills in mathematics for junior high schools students using a multiple-choice test with four options (Rosnawati, Kartowagiran, & Jailani, 2015, pp. 189–196).

Multiple-choice tests frequently studied are those of the NE and NSEE. Some of the problems related to the use of these two tests are the quality of the test and frauds frequently occur during test administrations. A study shows that, based on item response theory analyses, of the 40 items of the Mathematics NE for the junior high school, 28 are good and 12 are poor (Kartianom & Mardapi, 2017, p. 172). To look at the fraud practices during the administering of the national examination can be done from the NE integrity indexes.

In some regions, integrity indexes are found low, showing high fraud in the administration of the exams. This condition indicates that students of the primary and junior secondary schools are still fearful of the exams, although the results are not the only determinations for passing. The national exam, however, is used as a criterion for admission to the higher school level. For such, students give all kinds of efforts to get good results; one of which is by sharing answer keys. The multiple-choice system makes it possible for the test takers to exchange answers easily. This chance raises illegal cooperation among the test takers, which cause the test results to be invalid. Consequently, the exam results do not at all reflect the real competences of the students.

This problem needs a solution. One solution taken by the government is by giving out several parallel tests. Development of parallel tests takes different ways among subject matters in its method and rules. In the mathematics subject matter, item stems and options involve a lot of figures. Differences in the figures can have an impact on the levels of item difficulties. Even numbers and odd numbers give different difficulty levels. The choice of distractors also influences difficulty levels. In the development of the test packages for mathematics, therefore, must obey the rules.

In another angle, mathematics teachers are expected to prepare the students in approaching the national examination. In order to know the teachers' readiness to do it, research needs to be conducted. A study on the competence and readiness of mathematics teachers looked at the self-efficacy of mathematics teachers in Yogyakarta. The findings show that the self-efficacy of 43.07% of the teachers is at the low category, 55.47% at the medium category, and the rest 1.46% at the high category (Widdiharto, Kartowagiran, & Sugiman, 2017, pp. 69–75). These findings indicated that teachers' confidence in facing the NE was at the medium level. Probing further on the competence and readiness of teachers in approaching the NE and NSSE, it was necessary to know the teachers' competencies in developing test practices and try-outs for the NE. The purpose of the try-outs was to see each student's competence achievement to be used as a basis for improvement activities. It is, therefore, crucial that the test items developed by teachers be functional in showing the students' competences.

Another thing to be conducted is that which could minimize students' interaction in doing the test. This minimalization is done by developing several test packages. The packages should be parallel so that they would not raise a new problem. A parallel test must have

identical objective, difficulty level, and format so that the test will be the same, but the items will be different. If the packages have been able to minimize frauds but have different levels of difficulty, the results will not be valid either. It is, therefore, necessary that the development of the test packages consider the parallelism of the items that are developed by teachers through a variety of methods. Before testing the parallelism of the test packages, it is necessary to gather information concerning the methods used by the teachers to develop the test packages. This paper is to figure out how teachers develop parallel test items of the multiple-choice type in mathematics.

Method

The research employed a descriptive research approach to obtain information about the methods that the teachers used for developing the mathematics test packages in the school. The study used interviews and document reviews as the test techniques and questionnaires as the non-test technique for collecting pertinent data. Open-ended interviews were given to 22 mathematics teachers. Each teacher was given the freedom to provide information to the method he/she used in developing the test packages. Each teacher was allowed to have more than one response, depending on his/her experiences.

The research instrument used to gather data was an interview guide. It contained questions about the methods to be used by the teachers to develop the test packages and the reasons for selecting the methods. In order to obtain evidence that the teachers did use the packages, documents review was done. Besides finding that the packages were there, it was also used for finding results of the tests to the students.

The questionnaires were used to look at the procedures for developing the packages. They were used to know the steps the teachers employed in developing the packages from the formulation of the objectives, construction of the specification table, to the item validation of content and language. They were also used to obtain evidence on the consistency between the item development and the test development procedure. The questionnaires were completed by check and cross marks. A check mark was given if a teacher did the step in the test development, a cross mark when a teacher did not.

Findings and Discussion

Findings

The key findings of the study are that in developing mathematics test packages, teachers had applied five methods including (1) randomizing the item numbers; (2) randomizing the sequences of response options; (3) writing items using the same contexts but different figures; (4) using anchor items; and (5) writing different items based on the same specification table.

The majority of teachers up to 37 % (of 22 teachers) used the same contexts with different figures to construct test items (as seen in Figure 1). It was followed by 21 % that developed different test items from the same table of specification. Meanwhile, other proportions developed the same items in different item numbers, developed the same items with different orders for the options, and used anchor items.

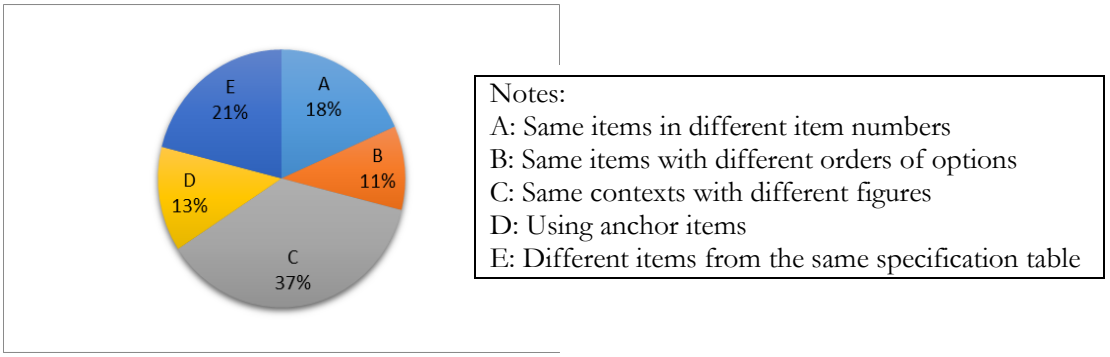


Figure 1. Methods of test package development

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

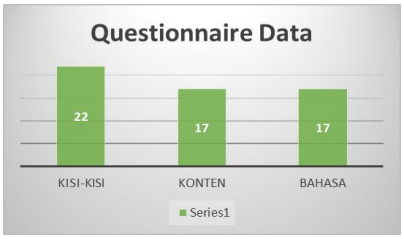


Figure 2. Data of instrument validation

Table 1. Randomization of the orders of response options

Package	Sample Item
Package 1 Sequencing from small to large numbers	Line gradient passing through point A(n, 3) and B(6, 2n) is 7. Value of n is … . A. 2 B. 3 C. 4 D. 5
Package 2 Sequencing from large to small numbers	Line gradient passing through point A(n, 3) and B(6, 2n) is 7. Value of n is … . A. 5 B. 4 C. 3 D. 2

Figure 2 presents a diagram of the results of questionnaires completed by 22 respondent teachers. It shows that all the teachers constructed the specification table before beginning to write the test items. Next, 17 teachers had their items validated in content and language by peer teachers. The rest five teachers did not have their items validated.

In developing test items, one should follow all steps set up in the procedure. After writing the items, teachers should have subjected them to peer validation by their colleagues as experts (Torres et al., 2011, p. 7).

Randomizing Item Numbers

From the interview, 18% of the teacher state they randomized items numbers to produce parallel items. Thus, the same test items were developed but were sequenced in different numbers. The difficulty levels and differentiating powers of the items were the same. The distractor functioning was the same too because identical distractors were used.

The method of randomizing item numbers is easy to use, does not take much time, and produces many test packages, as many as the test items. The interview reveals that some respondents commented that developing the items by changing the options order gave advantage to the students who got an item order that is the same with the content order. However, those who got items orders that are different from the content orders were put to a disadvantage because mathematics is built of axiomatic and deductive systems such that content sequences are highly compact.

Randomizing Sequences of Options

A total of 11% of respondents experienced randomizing the order of the response options. In developing multiple-choice test, randomizing the response options orders can minimize illegal interaction among the testees. The interview result reveals that randomizing the order of the options may result in two possibilities. First, if students find out that the options are different only in the orders, they can work out a way to interact with each other. In other words, this method still makes it possible for them to interact although they get different test packages. Second, if the students do not realize that the tests are different only in the options orders, they will not get advantage from their interaction. Thus, in this case,

the method functions well in minimizing frauds.

At least two test packages is needed in using this method, since sequencing can be done in two ways; from small to great or from great to small. An example of test package development by altering the options size is shown in Table 1. In Table 1, the stem in the two packages is the same, but the options order is different, although the options are the same. Test packages that have all the options in figures can only be developed in two different versions. If the response options are not in the form of figure, more packages can be obtained (Table 3). In this version, the stem and options are the same, but the options order is different. The number of packages that can be developed depends on the number of options. For example, a three-option item can be sequenced in several versions (Table 2).

Table 2. Randomization of response options

Package	Option Order
Package 1	A. P1
	B. P2
	C. P3
Package 2	A. P1
	B. P3
	C. P2
Package 3	A. P2
	B. P3
	C. P1
Package 4	A. P2
	B. P1
	C. P3
Package 5	A. P3
	B. P1
	C. P2
Package 6	A. P3
	B. P2
	C. P1

Table 3. Randomization of response options

Package	Item
Package 1	Line equation that passes the point (0. -2) and point (4. 1) is … . A. $y = \frac{3}{4}x - 2$ B. $y = \frac{4}{3}x - 2$ C. $y = \frac{3}{4}x + 2$ D. $y = \frac{4}{3}x + 2$
Package 2 Option A is exchanged with D and B with C.	Line equity that passes the point (0. -2) and point (4. 1) is … . A. $y = \frac{4}{3}x + 2$ B. $y = \frac{3}{4}x + 2$ C. $y = \frac{4}{3}x - 2$ D. $y = \frac{3}{4}x - 2$
Package 3 Option A is exchanged with C and B with D.	Line equation that passes the point (0. -2) and point (4. 1) is … . A. $y = \frac{3}{4}x + 2$ B. $y = \frac{4}{3}x + 2$ C. $y = \frac{3}{4}x - 2$ D. $y = \frac{4}{3}x - 2$
Package 4 Option A is exchanged with B and C with D.	Line equation that passes the point (0. -2) and point (4. 1) is … . A. $y = \frac{4}{3}x - 2$ B. $y = \frac{3}{4}x - 2$ C. $y = \frac{4}{3}x + 2$ D. $y = \frac{3}{4}x + 2$
Etc.	

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

From Table 2, it can be seen that six test packages can be developed from changing the orders of three response options. The number of test packages obtained by randomizing the orders of the options is n!, where n is a number of options. The number of packages can increase if a number of ways of item combination are impacted by certain items. Suppose Package 1 is an initial item; Packages 2 – 6 can be constructed in the way shown in Table 3. Package 7 can be constructed by exchanging options A and B on an even item and B and C on an odd item. This way of combining items will give a larger number of packages.

Table 3 presents an original item of two test packages with no order of options. Package 2 is obtained by changing options A and B in the initial item. The number of options influences the number of packages. Generally, the mathematics items for primary and junior secondary schools have four options, while senior secondary schools have five options.

The method of constructing test items by changing the orders of the options is intended to maintain item characteristics. Also, the distractors are also expected to function effectively. Numerous studies have been conducted that are related to the quality of multiple-choice tests. The studies commonly look into the quality of items in terms of levels of difficulty, differentiating powers, and distractor effectiveness. In addition to revealing information about test qualities, these studies also look into aspects that need to be improved to increase the quality of tests to be able to measure well.

Constructing Items Using the Same Context but Different Figures

In the study, 37% of the respondents constructed the test items using the same contexts but different figures. This method (see Table 4) results in two test packages that will be able to minimize the testees' interaction.

The teachers revealed that this method of test construction decreases the students' chance to cooperate. However, item construction using this method should be done carefully by paying full attention to the figures being used in each package. Even though the figures in each package are different, care must be taken in terms of even and odd figures since there are different perceptions of these figures between boys and girls (Wilkie & Bodenhausen, 2015, pp. 3–9). Besides, the size of the figures must also be taken into great account to make sure that the item difficulties are equal. Item difficulty levels influence discriminating powers; good items will we correctly answered by 30% to 80% of the testees (Kehoe, 1995a, p. 1). These percentages must be taken care of so that the test administration is minimized from frauds, and the results are fair to all the testees.

Table 4. Items with the same context but different figures

Package	Item
Package 1	A room with an air-conditioning of 3^{o}C. After the device is activated, the room temperature reduces 2^{o}C every 4 minutes. When the air-conditioner has been activated for 28 minutes, the room temperature will presently be … ^{o}C. A. –20^{o}C B. –15^{o}C C. –12^{o}C D. –11^{o}C
Package 2	An air-conditioning set is 5^{o}C. After it is activated, the temperature of the device reduces 4^{o}C every 8 minutes. When the air-conditioner has been activated for 32 minutes, its temperature will presently be … ^{o}C. A. 21 B. 16 C. –11 D. –59

Table 4 presents two test items into different test packages developed with the same context but different figures. Test item 1 uses figures that are relatively smaller than those in Package 2. The combination of even and odd numbers, however, is equal. Item 1 uses 3, and item 2 uses 5. These are two odd numbers with a small difference. Later on, Package 1 uses 4 while package 2 uses 8. Meanwhile, 32 and 28 are not far apart; both are two-digit and even figures.

Using Anchor Items

From the interviews, 13% of respondents developed the test packages using anchor items. Some studies have been done to obtain evidence for the functioning of anchor items. Studies show that the more anchor items used, the better the results are for the test equalization (Kartono, 2008, pp. 317–318). It means that anchor items function to equalize tests. One study increased the anchor items of a physics test up to 40%; the results show that items at the low, mid, and high difficulty levels are not yet equal (Abdullah, Mansyur, & Rosdiyanah, 2016, pp. 217–218). This inequality may be due to the fact that physics tests involve items with figures in them. The use of different figures in items will have an impact on the item difficulty levels. Even and odd figures also influence difficulty levels. Mathematics subject matter involves a lot of figures in its tests; and, thus, in using this method, developers must be accurate and careful to produce parallel tests.

Developing Items Using the Same Specification Table

Based on the results of the interviews, 21% of respondents constructed a test specification table and developed from it some different test packages. This mode of instrument development can be done in several ways, such as using various figures in the test items, making the same problem with different contexts, etc. This method of test development is effective in reducing frauds when the test is based on the teacher's narratives.

The two test items presented in Table 5 are developed from the same indicator, problem-solving in daily life using line arithmetic. The contexts and figures used in the items are different. In package 1, what is known is the first leg and amount of increase per year; while in package 2, what is known is the line from leg 1 to leg 3. The figures used in the two items are also different. The teacher needs to pay attention to these differences. The case is feared in which students can complete package 1 but not package 2 because of the different contexts. This condition may cause invalid testing so that the objective of the evaluation is not achieved. In order to prevent this from happening, it is suggested that teachers know and have information about parallel testing and the ways to develop parallel tests.

Table 5. Items constructed out of the same indicator

Type	Item
Package 1	Amount of sugar consumption by people in a village is 1,000 kg in 2013 and is always doubled each year. The total sugar consumption from 2013 to 2018 is … . A. 66,000 kg B. 65,000 kg C. 64,000 kg D. 63,000 kg E. 62,000 kg
Package 2	A scavenger collects trash plastic bottles. On the first day, he gets 2.5 kg, on the second day 3 kg, and on the third day 3.5 kg, and so forth following an arithmetic line system. If the plastic bottles are sold to a collector at Rp10,000.00/kg, in 15 days the scavenger earns …. A. Rp800,000.00 B. Rp900,000.00 C. Rp1,000,000.00 D. Rp1,200,000.00 E. Rp1,500,000.00

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

Discussions

The research findings show that 18% of the respondents stated that they developed test packages by randomizing the order of the item numbers believed to be able to produce parallel sets of items. This method had also been done in the entrance testing at Muhammadiyah University of Bengkulu. The randomization of item numbers used the Linear Congruent Method (LCM) computer software. This selection system ran effectively (Gunawan & Prabowo, 2017, pp. 144–151). The test consisted of 100 items scheduled for 90 minutes. One of the test items is numerical. This test item has identical characteristics as numerical items tested in the school mathematics so that the method of randomizing the item numbers is effective. One advantage of this method of developing parallel tests of the multiple-choice type is that it can produce test packages in a large number. The number of test packages will be the same as the number of test items. It is the combination of all items in the test. A simple illustration of a test with three items can be seen in Table 6.

Table 6. Randomization of test item numbers

Package	Item Number
Package 1	1, 2, 3
Package 2	1, 3, 2
Package 3	2, 3, 1
Package 4	2, 1, 3
Package 5	3, 2, 1
Package 6	3, 1, 2

A test with three items can be developed into six test packages. The number of the packages is the combination of all the test items; so, if a test has an n item, the number of the packages that can be developed is n. A test consisting of 40 items can be developed by randomization of the item numbers into 40! packages.

Findings show that 11% of the respondents developed the packages by reordering the response options. In 2016, a study investigated the influence of distractor revision upon item validity and reliability. The study found that it did (Ali, Carr, & Ruit, 2016, pp. 6–9). Some other studies reveal that the quality of an item is influenced by the quality of distractors. Another study found that the quality of distractors has an impact on the item's difficulty level (Tarrant & Ware, 2010, pp. 539–543). The number of distractors, on the other hand, does not impact the item quality (Royal & Dorman, 2018, pp. 3–5). In conclusion, by maintaining the parallelism of the distractors, parallel instrument packages can be obtained.

In the interviews with the teachers, it was found that they randomized the response option by using google doc. It was a computer application for on-line testing. In the process, the teacher input a test set through the application. Google doc. would automatically shuffle the response options of each item. When the students open the application to do the test, they will get items with different orders of the options. This application helped teachers in providing test packages by using one initial test set. This computer application can be used with, of course, the backing of the school facilities for on-line testing. One weakness, however, lies in the fact that the computer application did not sequence figures from small to large or from large of small. It becomes a violation of the rules for randomizing response options. The use of google doc application must consider the form of the options. It would be best used for options that do not use series orders such as sizes of figures.

The method of constructing test packages from the same table of the specification was claimed by 37% of the respondents. Conditions and considerations must be taken into account when developing test packages using this method. However, not all the rules were followed. The teachers merely considered the contexts to get parallel levels of difficulty. As can be seen in Package 1 and Package 2, the options consist of one correct answer and three distractors.

Determining the correct answer within the options was almost not a problem. The problem lies, however, on providing distractors that can function well. Instrument development must also consider the parallel functioning of the distractors because they also contribute to the quality of the item. Distractors were made to lead low students to select them so that the item can distinguish between

low-achieving and high-achieving students. Worse, it should not happen that low-achieving students choose the correct answer while high-achieving students choose the wrong options. In this case, distractors do not function well. Table 7 presents some possibilities to help distractors functioning.

Based on Table 7, the possibilities of students' errors can be used as a basis for selecting distractors effectively. The distractors in Package 2 are 21, 16, and -59. For the item in Package 1, if distractors are calculated in the same way as they are in Package 2, the values 19, 14, and -53 are obtained. The item sample of Package 1 in Table 4 shows that the distractors are -20, -15, and -12. It shows that there is no parallelism in selecting distractors so that the item parallelism is doubted. Students' inaccuracy in doing Package 2 makes them choose the wrong options or distractors. Students' error in doing Package 1, if there are no good distractors, will induce them to try to find the correct answers. It may produce unfairness among testees.

From the interviews results, it is known that 13% of the respondents used anchor items to develop the packages. Development of test packages using anchor items has been done for NSSE for primary, junior secondary, and senior secondary schools, in addition to the NE. For the school examination (NSSE), the teachers were involved in developing the test items. Some items are standardized by the government, and the other is developed by the teachers. This is the anchor-based development. The anchor items function to equalize one item among the others. It is expected that the test will be able to reveal students' competencies across regions using tests that are different but equal.

Based on the results of the interviews, 21% of respondents developed different test items from the same specification table. This method requires extra time when many packages are expected to be produced. Besides, the characteristics of the items produced may not be the same so that it needs the difficulty levels testing of the items in each package. In the practice of developing different items from the same specification table of the national level, it is never achieved to produce different items having the same difficulty level albeit being developed from the same table of specification (Herkusumo, 2011). This thought must be considered when developing different test packages based on the same specification table.

Table 7. Possibilities of errors made by testees

Type	Possibility of errors of Package 2	Possibility of errors of Package 1
Error 1	Option 21 is obtained from: $\frac{32\ \text{min}\ utes}{8\ \text{min}\ utes} = 4$, so decreasing 4 times. The decrease in temp. in 32 min: $4 \times 4^{o}C = 16°C$ Room temp after 32 min: $16°C + 5°C = 21°C$ (Room temp after)	 $\frac{28\ \text{min}\ utes}{4\ \text{min}\ utes} = 7$, so decreasing 7 times. The decrease in temp. in 28 min: $7 \times 2^{o}C = 14°C$ Room temp after 28 min: $14°C + 5°C = 19°C$ (Room temp after)
Error 2	Option 16 is obtained from: $\frac{32\ \text{min}\ utes}{8\ \text{min}\ utes} = 4$, so decreasing 4 times. The decrease in temp. in 32 min: $4 \times 4^{o}C = 16°C$ (Testee stops at temp drop)	 $\frac{28\ \text{min}\ utes}{4\ \text{min}\ utes} = 7$, so decreasing 7 times. The decrease in temp. in 28 min: $7 \times 2^{o}C = 14°C$ (Testee stops at temp drop)
Error 3	Option -59 is obtained from: $\frac{32}{4} = 8$, (error in selecting a number to calculate temp.). The decrease in temp. in 32 min: $8 \times 8 = 64°C$ $5°C - 64°C = -59°C$	 $\frac{28}{2} = 14$, (error in selecting a number to calculate temp.). The decrease in temp. in 28 min: $14 \times 4 = 56°C$ $3°C - 56°C = -53°C$

Conclusion and Suggestions

Conclusion

Teachers can use various methods of developing mathematics test packages by randomizing the item number, reordering the response options, using the same context with a different figure, using anchor items, and using the same table of specification. These methods are applied based on the respondents' logical thinking supported by analyses proposing that the test packages being developed are parallel. However, no theoretical bases have been used by the teachers in developing the tests. All the teachers used a specification table to develop tests while most of them had validated content and language.

Suggestions

Further research is needed to look at how the parallelism of the test packages can be developed among those five methods. Such research will be useful for the teachers to improve their theories and knowledge in developing parallel multiple-choice test items so that their evaluation of students is valid and reflect the real students' competences.

References

Abdullah, S., Mansyur, M., & Rosdiyanah, R. (2016). Pengaruh jumlah butir anchor terhadap hasil penyetaraan tes berdasarkan teori respon butir. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, *46*(2), 207–218. https://doi.org/10.21831/JK.V46I2.10935

Ali, S. H., Carr, P. A., & Ruit, K. G. (2016). Validity and reliability of scores obtained on multiple-choice questions: Why functioning distractors matter. *Journal of the Scholarship of Teaching and Learning*, *16*(1), 1–14. https://doi.org/10.14434/josotl.v16i1.19106

Gunawan, G., & Prabowo, D. A. (2017). Sistem ujian online seleksi penerimaan mahasiswa baru dengan pengacakan soal menggunakan Linear Congruent Method (Studi kasus di Universitas Muhammadiyah Bengkulu). *Jurnal Informatika Upgris*, *3*(2), 143–151. https://doi.org/10.26877/jiu.v3i2.1872

Herkusumo, A. P. (2011). Penyetaraan (equating) ujian akhir sekolah berstandar nasional (UASBN) dengan teori tes klasik. *Jurnal Pendidikan Dan Kebudayaan*, *17*(4), 455–471. https://doi.org/10.24832/jpnk.v17i4.41

Kartianom, K., & Mardapi, D. (2017). The utilization of junior high school mathematics national examination data: A conceptual error diagnosis. *REiD (Research and Evaluation in Education)*, *3*(2), 163–173. https://doi.org/10.21831/reid.v3i2.18120

Kartono, K. (2008). Penyetaraan tes model campuran butir dikotomus dan politomus pada tes prestasi belajar. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *12*(2), 302–320. https://doi.org/10.21831/pep.v12i2.1433

Kehoe, J. (1995a). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation*, *4*(10), 1–3.

Kehoe, J. (1995b). Writing multiple-choice test items. *ERIC/AE Digest Series EDO-TM-95-3*, *3*, 1–6.

Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.

Rasyid, H., & Mansur. (2008). *Penilaian hasil belajar*. Bandung: CV Wacana Prima.

Reynolds, C. R., Livingston, R. B., & Willson, V. L. (2009). *Measurement and assessment in education* (2nd ed.). Upper Saddle River, NJ: Pearson.

Rosnawati, R., Kartowagiran, B., & Jailani, J. (2015). A formative assessment model of critical thinking in mathematics learning in junior high school. *REiD (Research and Evaluation in Education)*, *1*(2), 186–198. https://doi.org/10.21831/reid.v1i2.6472

Royal, K., & Dorman, D. (2018). Comparing item performance on three- versus four-option multiple choice questions in a veterinary toxicology course.

Veterinary Sciences, *5*(2), 55. https://doi.org/10.3390/vetsci5020055

Stiggins, R. J., & Chappuis, J. (2012). *An introduction to student-involved assessment for learning*. Boston, MA: Pearson.

Tarrant, M., & Ware, J. (2010). A comparison of the psychometric properties of three- and four-option multiple-choice questions in nursing assessments. *Nurse Education Today*, *30*(6), 539–543. https://doi.org/10.1016/j.nedt.2009.11.002

Torres, C., Lopes, A. P., Babo, L., & Azevedo, J. (2011). Improving multiple-choice questions. *US-China Education Review*, *B*(1), 1–11.

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. https://doi.org/10.21831/reid.v3i1.13559

Wilkie, J. E. B., & Bodenhausen, G. V. (2015). The numerology of gender: Gendered perceptions of even and odd numbers. *Frontiers in Psychology*, *6*, 810. https://doi.org/10.3389/fpsyg.2015.00810

 Copyright © 2019, REiD (Research and Evaluation in Education), 5(1), 2019
ISSN 2460-6995

REiD (Research and Evaluation in Education), 5(1), 2019, 21-29
Available online at: http://journal.uny.ac.id/index.php/reid

Multiple intelligence assessment in teaching English for young learners

[1]Ernawati; [2]Hana Tsurayya; [*3]Abdul Rahman A. Ghani
[1,3]Department of Educational Research and Evaluation, Graduate School of Universitas Muhammadiyah Prof. Dr. Hamka
Jl. Warung Buncit Raya No. 17, Pancoran, South Jakarta 12790, Indonesia
[2]Early Childhood Education Center of Taman Anak Soleh As Salaam
Jl. Merapi Raya, Jatiwarna, Pondok Melati, Bekasi, West Java 17415, Indonesia
[*]Corresponding Author. E-mail: abdulrahman.ghani@yahoo.co.id

Submitted: 30 January 2019 | Revised: 01 March 2019 | Accepted: 11 March 2019

Abstract

There are many schools in Indonesia that provide English as one of their subjects. English has been taught from elementary schools, even in Kindergarten. Ironically, teaching English in most rural schools still uses a conventional method such as memorizing and translating. Many teachers cannot afford to provide well-designed, meaningful exercises for students to use on a one-to-one learning basis. As a result, students seem not having interest in learning English. Based on this reason, this study was conducted to identify students' intelligence through Multiple Intelligence Assessment to get effective approach in teaching English for young learners. The participants are an English teacher and students at an early childhood education center. This research focuses on presenting a deep description of the Multiple-Intelligence assessment to identify students' intelligence in order to get an effective way of teaching English for young learners. In collecting the data, three instruments were used: observation, interview, and document analysis. The findings of this study show that students have different interests and nature; some students love singing, some others enjoy drawing, and others like role-playing. Multiple-Intelligence Assessment helps the teachers to identify students' interests and bring them building some learning activities to attract them in learning English.

Keywords: *multiple intelligence, assessment, teaching English, young learners*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.23376.

Introduction

In recent years, Indonesian parents tend to choose an educational institution which provides English as one of their subjects. English becomes more popular because of its prestige as an International language.

The purpose of National Education is to develop students' potential to be faithful and devoted the Almighty, noble, healthy, knowledgeable, skilled, creative, independent, and responsible citizen of the democratic country (Law of Republic of Indonesia No. 20 of 2003). In order to reach the purposes, the government applies the 2013 curriculum which is based on knowledge, behavior, and ability competence cohesively to create productive, creative, and innovative students who are able to compete in globalization era.

During the last few years, the world of teaching witnessed the innovation of Teaching English for Young Learners. In Indonesia, as reported by Musthafa (2010), the government makes its own decision to put English as local content. This brings public awareness of learning English improved. English has been taught from elementary school, even in kindergarten.

According to Pinter (2006), language development starts well before children are

This is an open access article under the CC-BY-SA license.

able to say anything. Cameron (2005) also states in her book that children learn a second language better than adults. These opinions bring many kindergartens in Indonesia to offer English as one of their subjects. Indonesian parents engage their children to learn English at an early age.

In teaching English to young learners, teachers use many different approaches. All of the approaches have the same goal, to make an effective way in the learning process. Therefore, teachers build some activities to maximize students' potential in learning a language. Ironically, teaching English in most rural schools, still, uses a conventional method such as memorizing and translating. Many teachers cannot afford to provide well-designed, meaningful exercises for students to use on a one-to-one learning basis (Musthafa, 2010). As a result, students seem to have less interest in learning English. Maryanto (2005) on his research states the standard competency of Kindergarten teachers developed with Focus Group Discussion and Delphi Techniques have 50 indicators. One of these competencies is developing fun and interesting learning. Based on those reasons, teachers need to find a way in teaching English which provides interesting and enjoyable activities suited to children's interests and characteristics.

Students, as the object in the learning process, have a different nature. Some students enjoy singing, some others love drawing, while some of them like reading. According to Gardner (1983), all humans exhibit a range of intelligence. They are linguistic, logical-mathematical, musical, spatial, bodily-kinesthetic, interpersonal, intrapersonal, and natural. These differences indicate that children have a different way to enjoy the learning process. Pinter (2006) states that it is important for teachers to take into account that all children have stronger and weaker aspects of their multiple intelligences and preferred learning style.

Multiple Intelligences theory has been reported to be effective in teaching English to young learners. Hassan and Maluf (1999) who have conducted a study about the application of multiple intelligences in Lebanese kindergarten got the result that MI theory has successfully improved students' understanding in the learning process.

Multiple Intelligence assessment was conducted to identify the variety of students' intelligence. After knowing the interests of each student, the teacher will be easier to get an effective approach in teaching English for them. In detail, this study attempts to find out the activities and benefits of the implementation of Multiple Intelligence theory in teaching English for young learners.

This study is expected to contribute to the development of Multiple Intelligence Assessment in the learning process, especially in teaching English for young learners in Indonesia. In addition, it can give inspiration for teachers in order to make and create some activities that can help students to improve their ability in learning a language and maximize their potential in the learning process.

Teaching English for Young Learners

The term 'young learners' has been defined by Pinter (2006) as children who start their primary schools, either in kindergarten or elementary school. Wright (2001) states the specific age range of young learners, which is between age five to 12.

Children construct knowledge through other people, and through interaction with adults. Adults/teachers work actively with children in the Zone of Proximal Development (ZPD). Zone of Proximal Development (ZPD) is the difference between the child's capacity to solve problems on his own and his capacity to solve them with assistance. The adult's role is very important in a child's learning process. Like Vygotsky, Bruner focused on the importance of language in a child's cognitive development. He shows how the adult uses 'scaffolding' to guide a child's language learning through finely-tuned talk (Cameron, 2005).

Characteristics of Young Learners

Young learners or children have some typical characteristics. Piaget, as cited by Cameron (2005), states that a child is an active learner. They have a huge curiosity about learning something new. Children are often

more enthusiastic and lively. However, they also lose interest more quickly and are less able to keep themselves motivated on tasks they find difficult (Cameron, 2005). Further, Shin (2013) in her module entitled 'Teaching English to young learners' has classified the characteristics of young learners in learning English.

Assessment

Assessment is an integral part of educational processes Ghani (2008). According to Fenton (1996), assessment is the collection of relevant information that may be relied on for making decisions. In addition, Davies (2000) states that 'assessment for learning is ongoing, and requires deep involvement on the part of the learner in clarifying outcomes, monitoring on-going learning, collecting evidence and presenting evidence of learning to others.' She further points out that assessment which directly supports learning has five key characteristics: (1) learners are involved so a shared language and understanding of learning is developed, (2) learners self-assess and receive specific, descriptive feedback about the learning during the learning, (3) learners collect, organize, and communicate evidence of their learning with others, (4) instruction is adjusted in response to ongoing assessment information, and (5) a safe learning environment invites risk-taking, encourages learning from mistakes, enables focused goal setting, and supports thoughtful learning.

Assessment can be designed to measure a wide range of abilities. Assessment is designed to measure the ability of students' high order thinking which can be done by developing several instrument evaluation. These instruments can be used to measure the ability of students' high order thinking skill: multiple choice, essay, performance evaluation, and rubric.

Multiple Intelligences

Multiple Intelligences theory is genuinely introduced by Harvard psychologist named Howard Gardner. Regarding Gardner (1999), Multiple Intelligence theory consists of seven bits of intelligence, they are linguistics, logical-mathematic, musical, spatial, bodily-kinesthetic, interpersonal, and also intrapersonal. In 1999, Gardner, as cited by Armstrong (2009), added an eighth, natural intelligence. This theory has become a tool that educators around the world seize with enthusiasm (Hoerr, 2000).

Gardner, as cited by Veenema, Hetland, and Chalfen (1997), defines intelligence as an ability to solve problems or create products that are valued in at least one culture, whereas Helding (2009, p. 195) defines intelligence as a biopsychological potential. The specific explanation of Multiple Intelligence can be seen as follows:

The first one is linguistic or verbal intelligence; verbal intelligence involves the mastery of language. People with verbal intelligence tend to think in words and have highly developed auditory skills. They are frequently reading or writing.

The second is logical-mathematic intelligence; it consists of the ability to detect patterns, reason deductively, and think logically. These are usually the children who do well in the traditional classroom because they are able to follow the logical sequencing behind the teaching and are, therefore, able to conform to the role of model student.

The third one is spatial intelligence. This intelligence gives a person the ability to manipulate and create mental images in order to solve a problem. Spatial thinkers 'perceive the visual world accurately, to perform transformations and modifications upon one's initial perception' (Gardner, 1999, p. 173). People with this kind of intelligence tend to learn most readily from visual presentation such as movies, pictures, videos, and also demonstrations using models.

The fourth is bodily-kinesthetic intelligence. It entails the ability to understand the world through the body. These people can use their body in very expressive skilled ways for a distinct purpose.

The fifth is musical intelligence. It makes use of sound to the greatest extent of possible. Those with musical intelligence have a firm understanding of pitch, rhythm, and timbre.

The sixth one is interpersonal intelligence. It consists of the ability to understand,

perceive, as well as discriminate between people's mood, feelings, motives, and also intelligence.

The seventh is intrapersonal intelligence. Intrapersonal deals more with the individual itself. It is the ability to know oneself and to understand one's own working.

The last is natural intelligence. It involves the ability to understand nature's symbols, to respect the delicate balance that let us continue to live. They have a genuine appreciation of the aspects of nature and how they intertwine.

Multiple Intelligence Activities

In teaching English to young learners, teachers are expected to give their best efforts to maximize students' potential. Lash (2004) believes that in order to assist the children in getting the most from their learning experiences, the teacher must first identify the areas of intelligence in which each child excels. By doing this, the teacher will be able to understand the children's learning styles, and thus, know the best way to help the children integrate their experience into their body of knowledge.

Armstrong, as cited by Lash (2004, pp. 14–15), explains that there are some activities that can help students to maximize their potential based on their dominant intelligence, as follows: (1) Linguistic: Learners who fall into this category enjoy word games, drilling, creative writing, and reading for pleasure. They enjoy listening to stories being told aloud. (2) Logical-Mathematic: They enjoy playing strategy games like chess and checkers. They are willing to spend lots of time working on logic puzzles, such as Rubik's cube. They enjoy putting things into categories and using reason to work through problems. (3) Spatial: Art may be one of the activities in which spatially intelligent persons might like to spend lots of time. They enjoy jigsaw puzzles and other visual activities. (4) Bodily-Kines-thetic: They are good at competitive sports. They need to touch things in order to learn more about them. These individuals are good at mimicking people's gestures, mannerisms or behaviors. They enjoy messy activities like working with clay or finger painting. (5) Musical: This one seems pretty obvious. Musically gifted learners enjoy playing musical instruments, singing, or collecting CDs. They are sensitive to environmental sounds and respond strongly to different kinds of music. (6) Interpersonal: They have lots of friends and enjoy socializing with others in large and small groups. They enjoy playing group games. They enjoy teaching others and are seen as natural leaders. (7) Intrapersonal: They have a realistic sense of their strength and weaknesses. They react strongly when controversial topics are discussed. They have a sense of self-confidence. (8) Natural: Persons with high naturalist intelligence enjoy being in natural environments. Hiking and camping might be listed as their hobbies.

Hoerr (2000) in his book entitled 'Becoming a multiple intelligences school' writes a table about how to arrange and create some activities based on children's intelligence. This table is adapted from 'Succeeding with Multiple Intelligences', by the New City School faculty, 2000 (Hoerr, 2000).

Method

This research focuses on presenting a deep description of the Multiple Intelligence assessment used by the teacher in teaching English for young learners. For this reason, the researchers used a descriptive-qualitative method.

Qualitative research is a holistic approach that involves discovery. It is also described as an unfolding model that occurs in a natural setting that enables the researchers to develop a level of detail (Cresswell, 1994 quoted in Williams, 2011). Furthermore, Keegan (2009) explains qualitative as a research design that primarily concerned with meaning rather than measuring.

There are several characteristics of qualitative research, namely: (1) the focus of the research is 'quality', (2) the aim is description, findings, and understanding, (3) the settings are natural, (4) the sample is small and purposive, and (5) the data collection consists of the researchers as the main instrument, interview, and observation (Alwasilah, 2008, p. 92).

The aim of the descriptive method is to examine the current event or phenomenon of the research (Alwasilah, 2008). For this reason, the qualitative approach of the descriptive method is suitable to this research because it could be used to explain the detail of multiple intelligence activities used by the teacher in teaching English for young learners. In addition, this study employed the descriptive qualitative method because there is no treatment given during the observation. This study only observes the phenomena happened in the classroom; in detail, this study was set to investigate the implementation of multiple intelligence activities in teaching English for young learners.

The participants of this study are an English teacher and students at an Early Childhood Education Center of Taman Anak Sholeh As-Salaam, which is located in Bekasi, West Java, Indonesia. The observation was held in two months, from February to April 2018.

After collecting the data, the data of the study were then analyzed by several steps as proposed by Huberman and Miles, as cited by Basrowi and Suwandi (2008): First, data reduction; the researchers reduced unnecessary information from the data obtained through observation. In this process, the data from observation and interview were a transcript. Second, data analysis; the data from observation, interview, and document analysis were analyzed.

Findings and Discussion

Findings

The findings are from the data gained through observation and interview. This chapter consists of two main points: (1) multiple intelligence activities in teaching English for young learners in the Early Childhood Education Center being the object of the study, and (2) the benefits and challenges of using multiple intelligence activities.

Based on the observation, the response of students to several Multiple Intelligence activities are related to their interests. The data are displayed in Table 1.

An interview was conducted to find out the teacher's perspective of using multiple intelligence activities and the benefits of using the activities. According to the result of the interview on teacher's perception of the benefits of using multiple intelligence activities for young learners, the teacher mentioned several benefits of using these activities. First, the various activities can stimulate students in learning and they can be more active in the learning process. Since the activities were quite different from those in the conventional approach, the teacher admitted that students seem to enjoy the learning process through song, games, drilling, riddle, and so forth. Second, the learning process can be more effective because students understand the material easily by doing several activities. Third, it can motivate the students to learn English because the activities can cover their interests.

In spite of its benefits, some challenges are found in the use of multiple intelligence activities. The result of the interview shows that some students thought that the use of various activities was confusing. The students may be confused when they have to change from one activity to another.

The other challenge of using multiple intelligence activities was a limited time. The teacher said that sometimes the students need extra time to do the activity such as games and riddle. It means that using multiple intelligence activities needs a longer allocated time.

From the findings, it can be inferred that the use of Multiple Intelligence activities in teaching English for young learners has several benefits and challenges according to the teacher's perception. The advantages are that the activities can make students more active in the learning process, make the learning process more effective, and make students more motivated in learning English. Meanwhile, the challenges are that a long time is needed for the teacher to prepare the activities, and some media are needed in order to make the activities run well.

Table 1. Multiple intelligence assessment applied in the Early Childhood Education Center of Taman Anak Sholeh As-Salaam

Teacher's Activities	Students' Activities	Type of Multiple Intelligence:	Students' Responses	Assessment
Singing '*Kepala, Pundak, Lutut dan Kaki*' song with body gestures	Following the teacher doing the same activity	Musical Bodily-Kinesthetic	Most of students seem happy and enjoy the activity	*****
Showing animals pictures and mentioning the name of each animal	Looking at the pictures and repeating after the teacher to mention the name of the animals	Spatial Verbal-Linguistic	Most of students seem happy and enjoy the activity	*****
Playing the riddle game and giving clues for students to guess the correct answer	Listening to the clues and information given by the teacher	Logic-Mathematic Verbal-Linguistic	Some students look confused because they do not understand some words said by the teacher	***
Counting numbers in sequence and counting down the numbers	Following the teacher doing the same activity	Logic-Mathematic	Most of students can follow the teacher well when they are asked to count from 1-10, but they look a bit confused when the teacher asked them to count down	*****
Showing picture and playing the 'What color is it?' game	Looking at the visualization of the new word	Spatial Verbal-Linguistic	Most of students seem excited to say the various colors in English	*****
Playing 'Take the ball' game is the activity referring to bodily kinesthetic intelligence because the activities in it are related to physical movement	The students have to run quickly to get the right ball; the fastest group that took the right ball is the winner	Bodily-Kinesthetic Spatial	Most of students seem excited to participate in this game	*****
Playing 'Parts of body' game is an activity referring to interpersonal intelligence	Following the instruction by touching parts of body of their desk mate	Interpersonal	Some students seem uncomfortable when they have to interact with and touch their desk mate	***
Showing the picture of a fish to the students and giving personal connection by asking them about the fish. The teacher asked students about their previous knowledge of the material object	Giving opinions on the teacher's questions	Spatial Intrapersonal	Some students need more stimulus to speak up	***
Asking students to look outside the class and asked them about the weather	Looking outside the window, observing the weather, and giving some opinions	Natural	Most of students seem happy and enjoy the activity	*****

Note:
***: only some students participate actively in this activity
*****: successfully attract most of students to participate actively in learning process

Discussions

The teacher used drilling as the activity referring to linguistic intelligence. This activity was found in every meeting. Drilling is a strategy to improve pronunciation by imitating and repeating words, phrases, even whole utterances (Thornbury, 2006). In the first meeting, the teacher showed animal pictures to students and mentioned the name of the picture. After that, the teacher asked the students to repeat the word.

From the observation, it was discovered that the teacher taught a new word by saying it repeatedly in the drilling process. After saying the word out loud several times, the students finally remember the word. Therefore, this activity can be an effective way to teach children about foreign language. Students enjoy doing this activity, and saying the new words together with their friends make students get motivated to speak English in the classroom. However, students seem a bit confused when they have to repeat a long sentence. Teacher should know how to divide the sentence into a shorter part, especially when drilling a lyric from a song.

The teacher used riddle and counting numbers in a sequence as the activities referring to Logical-Mathematic Intelligence. Riddle was used by the teacher when teaching about an animal. The teacher prepared some animal pictures as the media. Before the teacher showed the animal pictures to students, the teacher did a riddle by giving some clues to students about the animal. The students have to listen to the clues and information given by the teacher; after that, the students have to analyze the information and clues from the teacher. In this activity, they need to think logically before answering the riddle. If the students have known the answer, they can guess it by saying out loud and give the answer to the teacher. And the last, the teacher will show the picture and mention the name of the picture. The logical mind can be stimulated anytime information is put into some kind of rational framework (Armstrong, 2009). In this activity, giving some clues and information about the animal is the stimulation of logical thinking.

Spatial intelligence has something to do with pictures - either the pictures in one's mind or the pictures in the external world, such as photos, movies, drawings, graphic symbols, ideographic language, and so forth (Armstrong, 2009). Showing picture and also games 'what color is it?' are the activities referring to spatial intelligence because these activities involve visualizing the objects and creating a mental image. Showing picture was found in the first, fourth, and fifth meeting; and games 'what color is it' was found in the sixth meeting.

The teacher used some pictures as the media to teach learning material. The pictures are related to the topic of each meeting. From the picture, students get the visualization of the new word. Wright (2001, p. 10) said that picture can play a key role in motivating students, conceptualizing the language they want to use, giving them a reference, and in helping the discipline of the activity.

Body answers and playing 'take the ball' game are the activities referring to bodily-kinesthetic intelligence because these activities are related to physical movement. Responding to the instruction with physical gesture was found in the third and fourth meeting, while 'take the ball' game was found in the sixth meeting.

Playing the game 'Parts of Body' is an activity referring to interpersonal intelligence. In this activity, the students are supposed to be working in pairs. The teacher gives the instruction first, and then the students follow the instruction by touching part of the body of their chair mate. The students have to listen to the teacher's instruction carefully.

Personal connection is an activity referring to intrapersonal intelligence. This activity was found in the first meeting. In the beginning, the teacher showed the fish picture to the students and gave personal connection by asking them about the fish. The teacher asked students about their previous knowledge of the material object.

Observing the weather is an activity referring to natural intelligence. This activity mostly happened at the beginning of the class before the teacher gives lesson material to the students. In this activity, the teacher asked the

students to look outside the class and asked them about the weather. The students seem happy when the teacher asked them to look outside the window and observe the weather.

Conclusion and Suggestions

Conclusion

This study was concerned with identifying students' intelligence by Multiple Intelligence (MI) assessment and applying MI activities in the process of teaching English for young learners. According to the findings and discussions, there are several multiple intelligence activities employed by the teacher in teaching English for young learners. These activities are categorized into eight groups based on the eight multiple intelligences.

It can also be concluded that the teacher has to know students' characteristics, interest and ability to create interesting and suitable activities for students. The variety of activities may improve students' attention and motivation in learning English.

Moreover, the result of the interview shows that there are several benefits in using MI activities in teaching English for young learners. However, in implementing MI activities in teaching English for young learners, the teacher found many challenges.

Suggestions

After concluding the analysis, the researchers would like to propose some suggestions related to the research conducted. Since this study involved merely only one teacher as a respondent, further study is suggested to involve more teachers as respondents.

Furthermore, implementing multiple intelligence activities in teaching English for young learners can be a recommendation to be used by kindergarten English teachers. Teachers should find many activities of teaching English which may cover students' interests and intelligence. To improve teachers' knowledge of multiple intelligence activities, the teachers can attend seminars, workshops, or training. In addition, teachers can get more information about multiple intelligence activities by reading books or searching on the internet.

References

Alwasilah, A. C. (2008). *Pokoknya kualitatif: Dasar–dasar merancang dan melakukan penelitian kualitatif*. Jakarta: PT.Dunia Pustaka Jaya.

Armstrong, T. (2009). *Multiple intelligences in the classroom*. Alexandria, VA: ASCD.

Basrowi, & Suwandi. (2008). *Memahami penelitian kualitatif*. Jakarta: Rineka Cipta.

Cameron, L. (2005). *Teaching languages to young learners*. Cambridge: Cambridge University Press.

Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International*, *37*(4), 346–355. https://doi.org/10.1080/1355800007500052955

Fenton, R. (1996). Performance assessment system development. *Alaska Educational Research Journal*, *2*(1), 13–22.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

Gardner, H. E. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY: Basic Books.

Ghani, A. R. A. (2008). Pengaruh tes formatif dan kemandirian belajar terhadap hasil belajar ekonomi siswa SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *12*(2), 162–176. https://doi.org/10.21831/pep.v12i2.1425

Hassan, K. El, & Maluf, G. (1999). An application of multiple intelligences in a Lebanese kindergarten. *Early Childhood Education Journal*, *27*(1), 13–20. https://doi.org/10.1023/A:1026063322136

Helding, L. (2009). Howard Gardner's theory of multiple intelligences. *Journal of Singing*, *66*(2), 193–199.

Hoerr, T. R. (2000). *Becoming a multiple intelligences school*. Alexandria, VA: Association for Supervision and Curriculum Development.

Keegan, S. M. (2009). *Qualitative research: Good decision making through understanding people,*

cultures and markets. London: Kogan Page.

Lash, M. D. (2004). Multiple intelligences and the search for creative teaching. *Paths of Learning*, *Autumn*(22), 13–15.

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

Maryanto, A. T. T. (2005). Pengembangan instrumen analisis kompetensi tutor Pendidikan Anak Usia Dini (PAUD). *Jurnal Penelitian Dan Evaluasi Pendidikan*, *7*(2), 241–252. https://doi.org/10.21831/pep.v7i2.2023

Musthafa, B. (2010). Teaching English to young learners in Indonesia: Essential requirements. *Educationist*, *4*(2), 120–125.

Pinter, A. (2006). *Teaching young language learners*. Oxford: Oxford University Press.

Shin, J. K. (2013). Teaching English to young learners. Baltimore County: English Language Center.

Thornbury, S. (2006). *How to teach speaking*. New York, NY: Longman.

Veenema, S., Hetland, L., & Chalfen, K. (Eds.). (1997). Multiple intelligences: The research perspective, a brief overview of the theory. In *The project zero classroom: Approaches to thinking and understanding*. Cambridge, MA: Harvard Graduate School of Education and Project Zero.

Williams, C. (2011). Research methods. *Journal of Business & Economics Research (JBER)*, *5*(3), 65–72. https://doi.org/10.19030/jber.v5i3.2532

Wright, A. (2001). *Pictures for language learning*. Cambridge: Cambridge University Press.

REiD (Research and Evaluation in Education), 5(1), 2019, 30-40
Available online at: http://journal.uny.ac.id/index.php/reid

Developing a problem-based local history module to improve the critical thinking ability of senior high school students

[*1]Novita Mujiyati; [2]Warto; [3]Leo Agung Sutimin
[1]Graduate School of Universitas Sebelas Maret
Jl. Ir. Sutami 36 A, Kota Surakarta, Jawa Tengah 57126, Indonesia
[2]Faculty of Cultural Sciences, Universitas Sebelas Maret
Jl. Ir. Sutami No. 36A, Jebres, Kota Surakarta, Jawa Tengah 57126, Indonesia
[3]Faculty of Teacher Training and Education, Universitas Sebelas Maret
Jl. Ir. Sutami No. 36A, Jebres, Kota Surakarta, Jawa Tengah 57126, Indonesia
[*]Corresponding Author. E-mail: novitamujiyati2@gmail.com

Submitted: 15 March 2017 | Revised: 11 April 2018 | Accepted: 06 May 2019

Abstract

This research aims at developing a problem-based local history module as learning media to improve the critical thinking ability of Senior High School students. This research employed a research and development (R&D) method referring to ADDIE development design. The procedure of designing ADDIE development included (1) analyzing to identify the learning media needed, (2) designing to organize the material of module corresponding to basic competency and indicator, (3) developing to obtain the finished product of module through validation and limited and large-scale trials, (4) implementing the module during learning process, and (5) evaluating to find out the effectiveness of module. The results of the research show that the problem-based local history module effectively improves the students' critical thinking ability. Through this module, the students are directed to understand a variety of aspects affecting a problem and relating it to the knowledge owned. The problems organized in the module are contemporary and related to the student environment. The process of connecting new knowledge to pre-existing knowledge enables the students to develop their own knowledge and to improve their critical thinking ability in solving the problems.

Keywords: *module development, problem-based local history, critical thinking*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.13334

Introduction

History learning at senior high school contributes to improving the cognitive, affective, and psychomotor abilities of students. Kochhar (2008, pp. 54–64) suggests that the values found in history learning are (1) scientific, (2) informative, (3) ethical, and (4) nationalism values. Meanwhile, according to Aman (2011, p. 57), History subject plays a strategic role in creating the nation's disposition and prestigious civilization, and in creating Indonesians who have nationalism and love to the homeland. History is the subject instilling knowledge, attitude, and values concerning the change and development process of Indonesian and world people from the past to the present (Sutimin & Wahyuni, 2013, pp. 55–57).

Thus, history learning is not only limited to knowledge or memory about many events which have occurred. History learning should enhance the students' critical thinking ability concerning the change, the development, and also the interrelationship between events. In addition, the students who have ever learned history should have a wise attitude and understand their identity. In the learning process, the teacher is the learning designer aiming to make the change toward

This is an open access article under the CC-BY-SA license.

the better one and the mastery of particular skill within the students. The interaction between teachers and students and environment condition should be designed to encourage the students to understand and to respond critically to what they are learning. As Illeris (2011, pp. 9–24) suggests, learning is a broad process that should pay attention to internal and external condition of students, so that functionality, sensibility, and also sociality constituting the competency of students will be created.

Considering the results of preliminary study, history learning at senior high schools has not arrived yet at the stage of improving the students' cognitive domain, particularly critical thinking ability. Supriatna (2011) says that history learning so far is dominated by collective memory so that the construction of history learning is required through contemporary issues discussed using historical perspectives. The discussion of the contemporary problems using historical perspective enables the students to construct their own knowledge and to think critically because it gives direct experience and involves the students' analysis process. Meanwhile, if the learning activity is dominated with collec-tive memory, such the process has not been stated as learning. It is because an action can be said as learning if it includes three criteria: (1) involving the change, (2) lasting for a long time, (3) occurring through experience (Schunk, 2012, pp. 5–6). Meanwhile, the construction of contemporary issues-oriented history learning proposed by Supriatna theoretically indeed yields the meaningful learning activity. It is because the students involved actively in the analysis process will construct their own knowledge and think critically.

However, it is noteworthy that contemporary issues-oriented learning will not construct the students' knowledge and improve their critical thinking ability if it is not designed and supported with appropriate learning media. Masek and Yamin (2011) suggest that theoretically, problem-based learning can improve critical thinking ability. However, empirically, improving students' critical thinking ability through problem-based learning takes time and is affected by such factors as age background, students' achievement, learning design, and media support. Therefore, to achieve the objective in applying problem-based learning, the attention should be paid to facilitators' role and learning media.

Considering the aforementioned phenomena, this research aims at developing learning media in the form of problem-based local history module and module using strategy in the learning process. Meanwhile, the objective of learning process using problem-based local history module is to improve the students' critical thinking ability, because in the history learning process, the role of media, particularly module concerning the local history and module using strategy, is very desirable. Module, according to Daryanto (2013, p. 9) is one of the teaching materials packaged comprehensively and systematically, containing a series of learning experiences planned and designed to help the students master the learning objectives specifically.

It means that a module can be used any time by the students and not always dependent on the teachers' role. However, the use of the module in the learning process should consider the appropriate learning method and strategy. Thus, the module containing local history materials and applying contemporary issues will effectively practice the students' critical thinking ability in the learning process if it is applied using an appropriate strategy. Meanwhile, when the module is used to learn independently, it will facilitate the students in understanding materials and also constructing their understanding because it has contained problem solving instruction. The contemporary issues existing in the module allow the students to conduct analysis based on historical perspectivesin order to improve their critical thinking ability.

Method

This research employed a research and development (R&D) referring to ADDIE development design. ADDIE, according to Suranto (2015, p. 154), is a model and design to create a learning product or a design to implement a program. Furthermore, according to Pribadi (2014, p. 23), ADDIE design includes five stages of activity: analysis, (2) de-

sign (3) development (4) implementation, and (5) evaluation.

The research subject was 150 students of Class X1 at SMAN 1 Trimurjo. The subject was chosen by purposive sampling technique. The data collecting techniques in this research were observation, interview, document analysis, and questionnaire. Observation was conducted to know the history learning process at SMAN 1 Trimurjo. Then, an interview with history teacher and students about the obstacles during the history learning process was conducted. The interview result shows the lack of local history module at SMA Negeri 1 Trimurjo. In order to follow up the problem, document analysis was done in the form of learning tools and document analysis for the development of local history module. After the module has been prepared and used in the learning process, the questionnaires were conducted to the teacher and students about the effectiveness of the module.

The processing and analysis techniques of the preliminary data (pre-survey) gained through observation and interviews were measured using qualitative methods. The process of document analysis in the form of learning tools and module development was obtained through literature study; then the development results were validated by the material experts and media experts through questionnaires. From the description of the questionnaire, the module was revised and tested. For the implementation phase of the module, the data obtained were analyzed quantitatively through T-test. In analyzing the quantitative results, the SPSS program was employed in this study.

Findings and Discussion

Learning Media Use

Based on the results of pre-survey in SMA Negeri Trimurjo, through both observation and interview with History Subject teachers and students, it can be seen that the use of textbook supporting media such as module is still minimal. The results of observation on the learning process indicate that the media, such as local history module has not been used yet. Meanwhile, the result of the interview with teacher and students shows that the understanding on local history is still very inadequate, even the teachers state that in the learning process so far, the discussion about local history is still very inadequate. It is because of limited learning sources and teaching materials addressing the local history.

The lack of history learning media cannot be ignored, because it is one source of information in the learning process. Rosyad (2003, p. 120) says that, methodologically, learning media aims (1) to help explain the subject matter the teacher delivers, (2) to help the students in learning activity, (3) to help the teacher dealing with space and time, (4) to give real experience to the students. Learning media is an instrument for delivering the teaching messages (Arsyad, 2010, p. 3). The types of learning media include: (1) audio using the role of sound in delivering the learning, (2) visual is the media using visual sense such as silent movie, photograph, picture, chart, map, and the similar, (3) audiovisual is the media using sound and picture elements (Anitah, 2010, pp. 7–69).

Concerning the arguments above, it can be seen that media is very desirable in the learning process; for that reason, the development of learning media should take the students' need and learning environment into account. Learning media serves to deliver information to the students concerning the essence of learning. Without media use, the learning will not attract the students' interest and will not motivate the students to understand the information delivered. Meanwhile, regarding the compatibility of media to the students' need and learning environment such as infrastructure and affordability aspect, the media developed should be used flexibly. It is because not all schools have adequate infrastructure to apply a variety of media. Considering this fact, the type of media developed in this research is visual media in the form of a problem-based local historical module to improve the students' critical thinking ability. Simply, the function of media in learning process is represented in Figure 1.

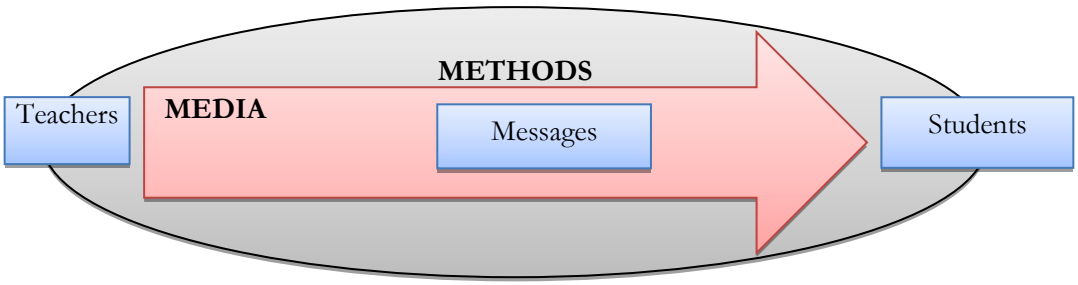


Figure 1. The function of media in learning process

Furthermore, local history is made the material in the module because it is the events familiar to and understandable to the students to create their knowledge. Wasino (2005, p. 1) says that historical learning should begin with the historical fact which is close (familiar) to the environment where the students live, and then the fact far away from the students' residence. Local history enables us to have an intimate relationship to the very local events which are not likely considered so far, but they are actually playing an important and valuable role in creating larger events (Abdullah, 1990, p. 19). Meanwhile, according to Aktekin (2010, pp. 86–105), historical learning utilizing local events will enrich knowledge and creates the students' identity.

Through local history, the students will feel learning about their own life. Something closely related to the students, despite smallness, will be more attractive than everything big but having no relationship. Also, in using the module, the appropriate strategy should be chosen to motivate them in finding and analyzing the source to find a resolution to the problems. The proper strategy in learning using the module is to use problem-based learning (PBL) in a group. Gallagher, et al., in Ward and Lee (2002, p. 17) say that Problem based Learning is the learning process, the learning starting point of which is based on the problem in real life. Thus, PBL will highly support the achievement of students' critical thinking ability.

Design of Problem-Based Local History Module

The design of module development is the process of organizing the problem-based local history material and adjusting the material with Basic Competency and Indicator. The local history material organized is the History of Lampung people's resistance in the 19th century. The Lampung people's resistance in the 19th century, according to the 2nd book of Lampung Province's Daily Council - Batch 45 (1995, pp. 84–94), occurred several times led by Raden Intan I (1808-1828), Raden Imba II (1828-1834), Bathin Mangunang (1817-1834), and Raden Intan II (1850-1856). Such resistance shows the social solidarity bond among Lampung people with different clans when dealing with the Dutch. The values of solidarity bond include the cooperation of many clans based on the feeling of responsibility, tolerance, and voluntariness to sacrifice in terms of material, effort, and also mind, for the sake of struggle.

Through analyzing the social solidarity values in the resistance of Lampung people in the 19th century, the students will be directed to find the resolution to some problems currently occurring in Lampung. The problem in Lampung is interethnic conflict due to the poor interaction between Lampung people and the incomers so that the social solidarity between the communities is not established. Meanwhile, the native Lampung people currently become more jostled in their own area. The problem of conflict in Lampung, according to Humaedi (2014, pp. 148–162) is derived from a narrow interaction due to inadequate common space leading to the failed cultural acculturation and the religion issue worsening the condition in Lampung. From this study, it can be found that poor social interaction leads to uncreated social solidarity between communities. Thus, there was no care, solidarity, common responsibility, and feeling of shared interest and fate.

It is the resolution obtained from historical analysis on the importance of social solidarity to Lampung people in the present that will be designed in the module. It is because to obtain the resolution, the students should understand their social environment and conduct historical analysis to recognize the values existing in Lampung people in the 19th century not existing in the present. Thus, through this understanding, the students can construct their knowledge and think critically. The compatibility of historical material concerning the Lampung people's Resistance in the 19th century to Basic Competency and indicator is presented in Table 1.

Table 1 shows that the material of Lampung people's resistance history in the 19th century will be included into the basic competency of the Indonesian Resistance Strategy against Western Colonialism before the 20th century and into the indicator of the Indonesian Struggle in many areas in fighting against Western Colonialism and Imperialism in Indonesia during the 20th century. The materials of Lampung people's resistance history in the 19th century which has not existed yet in the history textbooks will complement the students' knowledge to make them know that in their area, there had been a similar resistance with the other areas of Indonesia. The students' understanding of the Lampung people's resistance in the 19th century will facilitate the students to understand the resistance pattern in another area. In addition, the students will obtain a general description of people's resistance strategy before the 20th century more easily so that the basic competency and indicator of the learning can be achieved.

The Development of a Problem-Based Local History Module

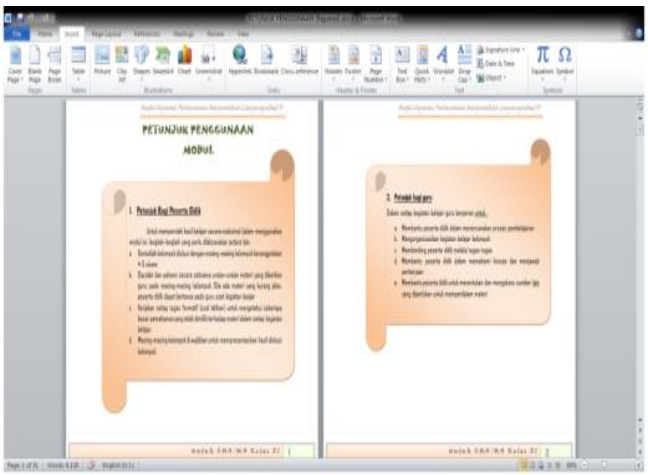
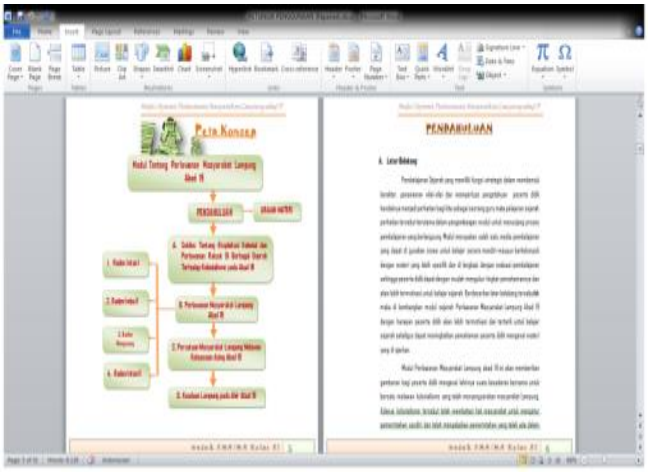
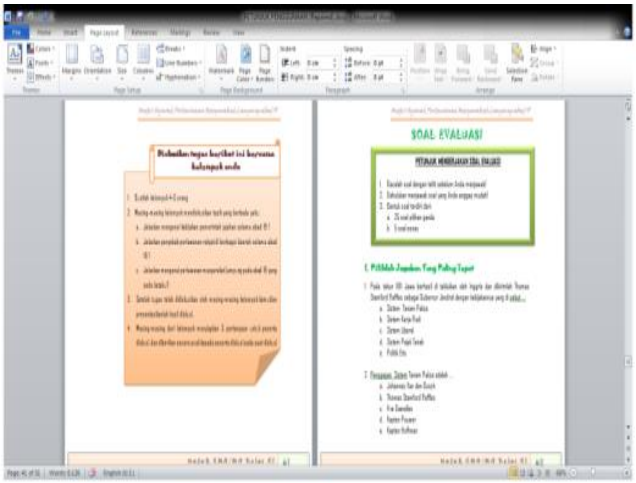
The development stage is the process of validating the initial draft of problem-based local history module by material and media expert, and limited and large-scale tryout with the students taken randomly to obtain the final draft module to be implemented in the learning process. A validator is selected based on the compatibility of expertise to the media to be validated. If the media to be validated is a problem-based local history module, the validator will be the one with media and material expertise concerning local history. Thus, the validators selected in this research were two media experts and two material experts. The initial draft module to be validated is presented in Table 2.

The result of the validator assessment of the initial draft on the title is (1) Material Expert I assesses with an average score of 4.34 the value is included in the 'good' category. The aspects assessed by the Material Expert I include the material feasibility, linguistic aspects, aspects of presentation, module effects on learning strategies and aspects of the overall look. (2) Material Expert II provides an assessment with an average score of 4.14, with a good category. The aspects assessed by the Material Expert II are the same as those assessed by the Material Expert I. This is conducted to provide a comparison of opinion between the two Experts of Materials to obtain information about the module material that is completely valid. (3) Media Expert I gives an overall average score of 4.54, which belongs to a very good category. (4) Media Expert II gives an average score of 4.5, which is also included in the category of very good.

Table 1. The compatibility of history material concerning the Lampung people's resistance in the 19th century to basic competency and indicator

Local History Material	History of Lampung People's Resistance in the 19th Century
Basic Competency	Analyzing the Indonesian Resistance Strategy against Western Colonialism before the 20th century.
Indicators	1. Explaining the western Colonialism and Imperialism process in Indonesia. 2. Explaining the impact of western Colonialism and Imperialism on Indonesia. 3. Analyzing the Indonesian struggle in many areas in fighting against Western Colonialism and Imperialism in Indonesia during the 20th century.

Table 2. The initial draft module to be validated

Initial Draft Module	Note
	Cover of module
	Module manual for teachers and students. Standard Competency, Basic Competency, and indicators, as well as learning objectives and benefits
	Concept Map, Introduction, and Material
	Summary, Evaluation Item and References

The aspects assessed by Media Expert I and Media Expert II are presentation aspects of including the attractiveness of the drawing, page design, cover design, font type, color selection, inter-page transition sustainability, as well as legibility. From these assessments, the local problembased historical module can be used for further tests with revisions according to the validator's suggestion.

The revisions that need to be done in the module are as follows: (1) The first suggestion from Media Expert I is that the module developed should use a learning model and show model steps. The second suggestion is a compiled module that must be aligned with the module-making guidelines. Besides, the selection of the colors used in the module should be more considered. Considering that the module users are all students, the colors used should also be general. (2) Media Expert II gives suggestions concerning the timing of implementation. According to Media Expert II, the module of Lampung people's resistance in the 19th century based on problem-based learning (PBL) would be better if it is implemented in two or three meetings, so that the historical values in the module can be comprehensively understood by the learners. (3) Material Expert I comments the number of years that do not yet exist, other than that the writing of the material must be based on a common event before heading to a more specific event of the 19th century Lampung society's resistance. (4) Material Expert II gives suggestions on the language use in the module which should be more communicative.

After the problem-based local history module has been improved based on the validators' suggestions, the next step is conducting a limited trial and extensive testing. The limited trial in this study includes two stages: a small group trial and large group trial. The result of the small group trial is a well-developed module with an average score of 4.2. From the small group trial process, the module was improved based on the advice from the students, as for the student's suggestion is that the writing on the module needs to be simplified to make it easier to understand. In addition, the tasks assigned in the module are too numerous, so that it needs to be reduced.

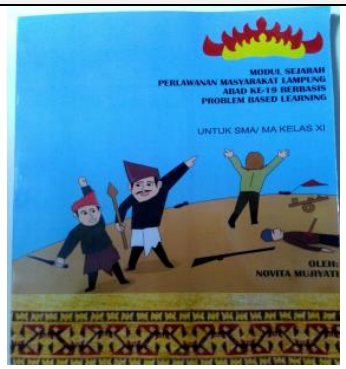
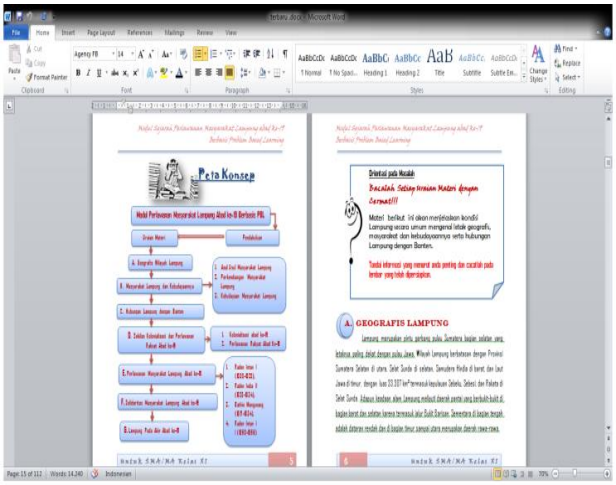
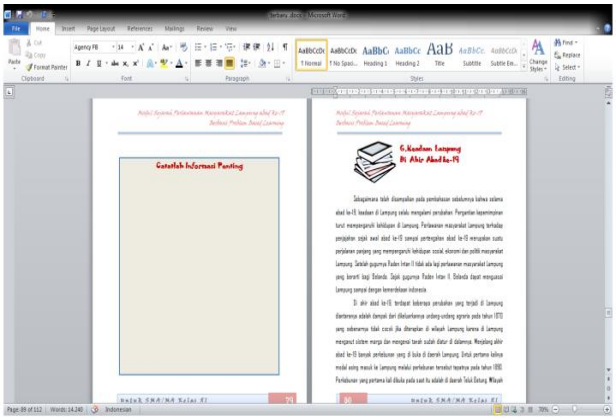
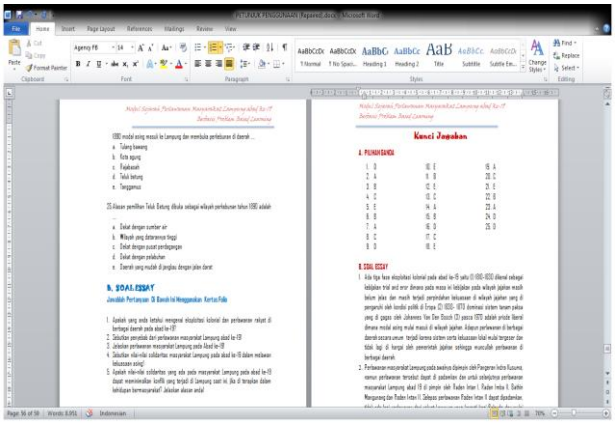
After the module is revised according to the suggestions, then, the large group trial was conducted. The result of the large-group trial is that the module was developed well with the mean score of 4.2 and stated as belonging to a good category with revision and ready to be used in a broad trial/tryout. In the broad tryout, the total score of 4.4 is obtained, belonging to a good category without revision. Thus, considering the assessment and approval from the history teacher for the 11th graders, the final draft module is obtained and stated as feasible to be used in the implementation stage. The final draft of problem-based local history module is presented in Table 3.

The Implementation of Module to Improve the Students' Critical Thinking Ability

The module implementation stage in the learning process is conducted in an experimental class through group work with problem-based learning (PBL) model. The PBL model, according to Barrow (in Huda, 2014, p. 71), is the one obtained through the process toward the understanding of a problem's resolution. The PBL model combines some disciplines in solving the problem (Anitah, 2009, p. 70). The learning can create high-order thinking ability and improve students' critical thinking ability (Sani, 2015, p. 127).

The module application using problem-based learning (PBL) model in a group will lead the students to construct the meaning of the learning process and to improve their critical thinking ability in solving the problem. In history learning, the problem-solving process is used to interpret, evaluate, and criticize historical document (Gredler, 2011, p. 308). In interpreting, evaluating, and criticizing processes, the students are required to have many analyzed sources to support an argument. It is in this process that the students' critical thinking ability is constructed. Slavin (2011) states that the teaching of critically thinking ability can be effective when the learning process emphasizes on the reason for an argument rather than on the correct answer. In addition, there is a process of connecting to the common topic to the students, so that various perspectives will be obtained allowing for the free discussion between students.

Table 3. The final draft of problem-based local history module

Content of Final Draft Module	Notes
	Cover design
	Concept map to facilitate the students understanding the material interrelationship. Learning materials and procedures to direct the students find the resolution of problem.
	Confirmation sheet to ensure that the students understand any material and contribute to learning activity. From the confirmation sheet, the teacher will find out the extent to which the students understand the learning material.
	Evaluation Items, Answer Keys, References

From the explanation, it can be seen clearly that the problem-based local history is the supplementary textbooks, supporting the students to explore their knowledge. The students will understand more easily the relationship between events from the narrow to the broader scopes. Meanwhile, the discussion process in finding the solution to the problem in the module leads the students to enter into the zone of proximal development (ZPD). Zone of proximal development (ZPD), according to Rezaee and Azizi (2012, pp. 51–57), can be achieved when there is a gap between the students' ability and the knowledge level achieved. In this stage, teachers should conduct scaffolding and discussion activity with their group to achieve the students' critical thinking ability maximally to solve the problems. The procedures for designing the discussion activity in the problem-based learning model are presented in Figure 2.

The effectiveness of problem-based local historical module can be seen through the comparison of pretest and posttest result of evaluation in control class and experiment class. Before the treatment, the two classes had approximately the same average value of 64, while after the treatment (learning using modules), the experimental class using the module got a higher value of 82. Meanwhile, the control class that did not use the module got the value of 70. The effectiveness test on the module was conducted in the 11th grade of Social Science 2 as the experiment class and the 11th Grade of Social Science 3 as the control class in SMA Negeri 1 Trimurjo. The test conducted in the two classes includes learning achievement test and questionnaire for students' social solidarity. Next, the result of the test was calculated in SPSS 21.

Considering the result of the data processed in SPSS 21, it is found that the normality test shows that the two classes obtain significant value > 0.05, meaning that they are distributed normally. The result of effectiveness test on the students' critical thinking ability shows the F value = 0.084 at the significance level of 0.773 > 0.05 so that there is a variance equation. Meanwhile, the result of t-test shows that tstatistic¬ value of 3.303 is at the significance level of 0.001< 0.05 so that H_0 is not supported, the mean score of experiment and control classes is not the same after the treatment. There is a positive significant increase in critical thinking ability in the class using a problem-based local history module.

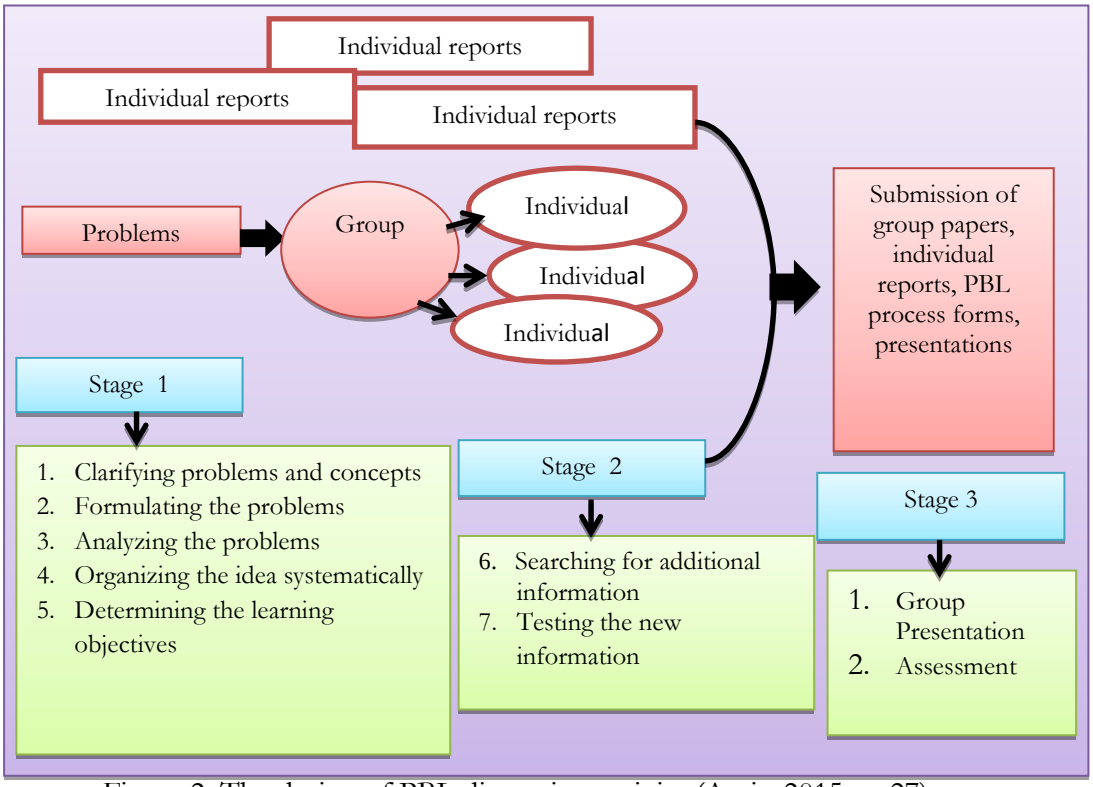


Figure 2. The design of PBL discussion activity (Amir, 2015, p. 27)

The result of the research is relevant to the finding of research conducted by Suarsana and Mahayukti (2013) on the development of problem solving-oriented e-module. This study found that (1) e-module developed is high in quality, but still needs improvement, (2) the use of e-module can improve the students' critical thinking skill, and (c) the students' respond to e-module use in lecturing is very positive. Furthermore, Styron (2014) suggests that in-group learning using the problem-based learning model effectively improve the students' critical thinking ability. The learning approach with critical thinking and building students' knowledge considerably contribute to the improvement of the students' achievement (Lunenburg, 2011, pp. 1–8). Considering some relevant studies in this research, and based on the effectiveness test conducted, it can be seen that the problem-based local history module effectively improves the students' critical thinking ability.

Conclusion

The development of problem-based local historical module using ADDIE design and applied through a problem-based learning model can effectively improve the students' critical thinking ability. Through such model, the students understand in-depth the relationship between events and expand their understanding of historical analysis to find the resolution to social problems existing in the students' environment. In addition, this research proves that the development of problem-based local historical module can be used as the supplement learning source for textbook.

This research shows that the effectiveness of the learning process is not only based on the learning model applied by the teacher. Meanwhile, the availability of teaching materials in the form of local-history modules that apply a problem-based learning model also becomes support in improving student's skills in critical thinking.

References

Abdullah, T. (1990). *Sejarah lokal di Indonesia*. Yogyakarta: Gajah Mada University Press.

Aktekin, S. (2010). The place and importance of local history in the secondary history education. *Journal of Theory and Practice in Education*, *6*(1), 86–105.

Aman, A. (2011). *Model evaluasi pembelajaran sejarah*. Yogyakarta: Ombak.

Amir, M. T. (2015). *Inovasi pendidikan melalui problem based learning: Bagaimana pendidik memberdayakan pemelajar di era pendidikan*. Jakarta: Kencana.

Anitah, S. (2009). *Teknologi pendidikan*. Solo: Inti Media Surakarta.

Anitah, S. (2010). *Media pembelajaran*. Solo: UNS Press.

Arsyad, A. (2010). *Media pembelajaran*. Jakarta: PT. Raja Grafindo Persada.

Daryanto. (2013). *Menyusun modul (Bahan ajar untuk persiapan guru dalam mengajar)*. Yogyakarta: Gava Media.

Gredler, M. E. (2011). *Learning and instruction - Teori dan aplikasi*. (Tri Wibowo B. S., Trans.). Jakarta: Kencana Prenada Media Group.

Huda, M. (2014). *Model-model pembelajaran dan pengajaran isu-isu metodis dan paradigmatis*. Yogyakarta: Pustaka Pelajar.

Humaedi, M. A. (2014). Kegagalan akulturasi budaya dan isu agama dalam konflik Lampung. *Analisa*, *21*(2), 149–162. https://doi.org/10.18784/analisa.v21i02.11

Illeris, K. (2011). *Contemporary theories of learning: Teori-teori pembelajaran kontemporer*. Bandung: Nusa Media.

Kochhar, S. K. (2008). *Pembelajaran sejarah (Teaching of history)*. Jakarta: PT. Gramedia Widiasarana Indonesia.

Lampung Province's Daily Council - Batch 45. (1995). *Sejarah perkembangan pemerintahan di Lampung* (Book II). Bandar Lampung: Badan Penggerak Pembina Potens iAngkatan 45.

Lunenburg, F. C. (2011). Critical thinking and constructivism techniques for improving student achievement.

National Forum of Teacher Education Journal, *21*(3), 1–9.

Masek, A., & Yamin, S. (2011). The effect of problem based learning on critical thinking ability: A theoretical and empirical review. *International Review of Social Sciences and Humanities*, *2*(1), 215–221.

Pribadi, B. A. (2014). *Desain dan pengembangan program pelatihan berbasis kompetensi (Implementasi model ADDIE)*. Jakarta: Prenada Media Group.

Rezaee, A. A., & Azizi, Z. (2012). The role of zone of proximal development in the students' learning of English adverbs. *Journal of Language Teaching and Research*, *3*(1), 51–57. https://doi.org/10.4304/jltr.3.1.51-57

Rosyad, A. (2003). *Teori belajar dan pembelajaran*. Jakarta: PT. Rineka Cipta.

Sani, A. R. (2015). *Pembelajaran saintifik untuk implementasi kurikulum 2013*. Jakarta: Bumi Aksara.

Schunk, D. H. (2012). *Learning theories: An educational perspective*. (E. Hamidah & R. Fajar, Trans.) (6th ed.). Yogyakarta: Pustaka Pelajar.

Slavin, R. E. (2011). *Psikologi pendidikan: Teori dan praktik*. (M. Samosir, Trans.). Jakarta: Indeks.

Styron, R. A. (2014). Critical thinking and collaboration: A strategy to enhance student learning. *Journal of Systemics, Cybernetics and Informatics*, *12*(7), 25–30. Retrieved from http://www.iiisci.org/journal/sci/FullText.asp?var=&id=EI597JP12

Suarsana, I. M., & Mahayukti, G. A. (2013). Pengembangan e-modul berorientasi pemecahan masalah untuk meningkatkan keterampilan berpikir kritis mahasiswa. *Jurnal Pendidikan Indonesia*, *2*(2), 264–275. Retrieved from https://ejournal.undiksha.ac.id/index.php/JPI/article/view/2171/1887

Supriatna, N. (2011). Konstruksi pembelajaran sejarah yang berorentasi pada masalah kontemporer pembangunan. *MIMBAR: Jurnal Sosial Dan Pembangunan*, *27*(1), 21–30. https://doi.org/10.29313/mimbar.v27i1.308

Suranto. (2015). *Teori belajar & pembelajaran kontemporer*. Yogyakarta: LaksBang PRESSindo.

Sutimin, L. A., & Wahyuni, S. (2013). *Perencanaan pembelajaran sejarah*. Yogyakarta: Ombak.

Ward, J. D., & Lee, C. L. (2002). A review of problem-based learning. *Journal of Family and Consumer Sciences Education*, *20*(1), 16–26.

Wasino, W. (2005). Sejarah lokal dan pengajaran sejarah di sekolah. *Jurnal Paramita*, *15*(1).

REiD (Research and Evaluation in Education), 5(1), 2019, 41-53
Available online at: http://journal.uny.ac.id/index.php/reid

An assessment instrument of mind map product to assess students' creative thinking skill

[*1]Regina Ratih Fardhila; [2]Edi Istiyono
[1,2]Department of Educational Research and Evaluation,
Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Karangmalang, Depok, Sleman, D. I. Yogyakarta 55281, Indonesia
[*]Corresponding Author. E-mail: reginaratihfardhila.2017@student.uny.ac.id

Submitted: 21 December 2018 | Revised: 22 April 2019 | Accepted: 06 May 2019

Abstract

Education is now more demanding for students to think more at a higher and more complex level. Mind map also builds creativity because it is more flexible in expressing ideas. The result of need assessment was the teachers cannot put on equation about creative thinking in test because the input of heterogeneous students, but they carried out project and practicum assessments, which required relatively much time, place and equipment. Therefore, the researchers develop instruments for assessing students' creative thinking skills at the high-school level trough mind map on valid and reliable biology subjects so that teachers can use them as a more efficient assessment tool. This research employed research and development (R&D) method; adapting ADDIE (Analyze, Design, Development, Implementation, and Evaluation) model. Content validity was achieved using the Lawshe method; CVR formula, while the construct validity was achieved through exploratory factor analysis with the SPSS program and then confirm with confirmatory factor analysis using Lisrel program. It is found out that the construct validity was proven by the significant value of MSA and Loading Factor value. Suitability of the construct model with the data was indicate with p-value= 0.68 (≥0.05); Root Mean Square Error Approximation = 0.00 (≤0.08); and Goodness of Fit Index= 0.94 (>0.90). The reliability of the instrument is 0.774, which means reliable. The assessment instrument of mind map product to assess creative thinking skill of students at the high school level in biology subjects has ten fit items in four factors; Flexibility, Elaboration, Originality, and Fluency.

Keywords: *product assessment, mind map, creative thinking skill*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.22525

Introduction

Education is now demanding students to think at a higher and more complex level, which is a high level of thinking on Bloom's taxonomy, namely analysis, evaluation, and creation. The top three levels of Bloom's taxonomy are called High Order Thinking Skills (HOTS). It is in line with the general objectives of 2013 curriculum which is to prepare Indonesian people to have the ability to live as individuals and citizens who are faithful, productive, creative, innovative and affective and able to contribute to the life of the world, nation, state, and world civilization.

The ability to think creatively becomes one of high-level thinking (HOTS). Creative thinking is generally regarded as gathering information to produce new understandings, concepts, or ideas (Moore, 2015, p. 380). The ability to think creatively creates a creative generation that has the potential to solve social problems and environmental issues (Yusnaeni, Corebima, Susilo, & Zubaidah, 2017). Creative thinking will produce fluency, flexibility, and originality (Buzan, 2011). Fluency in question will show speed in forming new ideas. Flexibility will indicate a person's ability to consider ideas and reverse the ideas that previously existed. Originality will form

This is an open access article under the CC-BY-SA license.

unique ideas. A creative person can think fluently, flexibly, has flexibility of thinking and originality, and also elaboration (DePorter & Henarcki, 2010, p. 292).

Learning assessment has an essential role in generating students' creative thinking skills because in general, a person's ability will be seen if there is an assessment activity aimed at assessing that ability. More specifically, the assessment or assessment process in education includes collecting evidence about students' learning achievements both through tests, observations, and also the work of the students.

Creative thinking is closely related to the creativity needed in biology learning because in biology learning, a student not only learns concepts, laws, principles, and facts but also learns about knowledge in the form of work craze, obtaining information on how science works and thinking skills (Herlina & Qurbaniah, 2017). Biology is one of the science aspects, so biology learning is demanded to make students have creativity in science process skill (Putri, Paidi, & Subali, 2016). Previous research from Putri et al. (2016) also stated that there was a relationship between divergent thinking ability of science process biology aspect of an elementary student, where divergent thinking ability is one of the creativity indicators. According to the creative ability model of Guilford (1950), someone creative has the characteristics of fluent thinking ability, flexible thinking ability (flexibility), original ability to think (originality) and elaboration ability.

Preliminary research data was obtained through interviews with four biology teachers in three different schools, namely Purbalingga 1 State High School, Purbalingga 2 State High School, and Purbalingga 1 Muhammadiyah High School. The four stated that they had not been able to assess students' creative thinking ability by conducting tests, both daily tests, and end semester assessments. They assess the ability to think creatively through projects and practicum, which are carried out 1-4 times in one semester, and of course, require a lot of time and equipment.

The results of observations on the grids and daily repetition questions as well as the final semester assessment of biology lessons used by the three schools showed that there were no questions that measured the ability to think creatively. Three teachers revealed the absence of a problem that measures creative thinking ability because the input of heterogeneous students is due, since the enactment of students admission by using an official letter stating that a student comes from a financially-dependent family, in the academic year of 2017/2018. Further, student admission in the academic year of 2018/2019 follows the Regulation of Minister of Education and Culture, No. 14 of 2018, on Student Admissions, which states that the zoning system of the distance between residence and school is the main criterion. With the existence of these regulations, the input of the students is increasingly heterogeneous.

From the result of need assessment, it is known that teacher needs a new way to assess creative thinking skill. Teachers need an assessment that more efficient; less time and equipment, and also can assess that heterogeneous student.

Mind maps can also be called thinking maps, which are pattern languages that allow all learners to have a language of visual-verbal cognition, to enable a deeper capacity to see, change, imagine, and improve their thinking skills (Hyerle, 2012, p. 1). Mind maps can also be called concept maps. Concept maps are visual presentations of concept connections and hierarchical organizations (Santrock, 2014, p. 5). Mind maps can be used as a tool to determine not only what has been studied or known by someone, but also how someone thinks about the information. The assessment using mind maps allows us to know how students think in receiving and processing information in learning.

Mind map also builds creativity because it is more flexible in expressing ideas. Presenting all relevant or relevant topics in the same mind map, detailing and connecting them through images, symbols, and warmth, can improve memory, it is easier to understand, and more effective in time and increase productivity (Murley, 2007). Another advantage of mind maps is that mind maps will provide a comprehensive view of the subject matter or

a large area, allowing us to plan routes or make choices, gather large amounts of data somewhere, encourage problem-solving by letting us see the path of creative breakthroughs, and fun to see, read, digest and remember (Putri, 2016).

Several previous research concerning the relationship between mind maps and the ability to think creatively has been carried out. A research which is conducted by Sari and Jarnawi (2008) states that learning through creative mind maps can improve creativity and student connections to be better. This research describes how creative mind map affects the students' creativity with pre-test and post-test control group design. Another research conducted by Putri (2016) states that mind maps can be used as an evaluation tool for students' creative thinking abilities after learning by mind map concepts in experimental classes. Other research on mind maps is carried out by Çoban and Tokatlı (2017), which states that mind mapping techniques affect a person's creativity, especially in music lessons. This is experimental research. Three of those previous researches have not developed an instrument to measure creative thinking skill.

There are also some previous researches that have developed an instrument to measure creative thinking skill. Research by Hartati, Asrial, and Ernawati (2017) develop an instrument to measure students' creative thinking skill in chemistry subject; the instrument is a subjective test instrument. It uses a development model proposed by Supardi, Haryanto, and Suhendri (2014). The construct validity was not using Exploratory Factor Analysis (EFA), and the content validity was not using Lawshe (CVR) either Aiken's method. Another research from Kristiani, Mayasari, and Kurniadi (2017) concludes that instrument developed is a subjective test instrument in the physics subject. It used a 4-D model. This research used Lawshe method for content validity, but it does not use EFA for construct validity. From the two previous researches which developed an instrument to measure creative thinking skill, there were have not developed a non-test instrument which used Lawshe method for content validity also EFA for construct validity. Whereas, to compile an instrument, validity of content and construct based on the existing theory is needed to produce a truly valid instrument. Both of them also have not counted the instrument's estimate of reliability. Therefore, in this research, EFA was used because the researchers want to know that the indicators made are included in the correct factors or latent variables.

Biology is one of the sciences in which, in its learning process, it needs to be more than just knowing, but also having to understand. In addition, finding out how students think in processing information also needs to be considered.

Mind map explanation is the basis that with mind maps, students' creative thinking skills can be assessed. The use of mind maps is expected to be able to see students' creative thinking abilities in processing biology learning material. Trough mind map, teacher can assess the students' creative thinking skill.

The previous explanation requires assessment tools that do not need relatively much time and equipment and can measure the creative thinking skills of heterogeneous students. This consideration becomes the basis of the researchers to develop instruments to assess students 'creative thinking abilities through mind maps on valid and reliable biology subjects so that they can be used by educators or teachers as a more efficient assessment tool for students' creative thinking skills.

Thus, the objective of this research are: (1) to develop an instrument to measure student's creative thinking in biology subject, and (2) to find out the validity and reliability of the instrument to measure student's creative thinking skill in biology subject.

This research is hoped to be a reference for a teacher who wants to develop an assessment instrument. The instrument produced in this research can be used by the teacher to assess students' creative thinking skill, especially in biology subject.

Method

The method used in this study is research and development (R & D) or better known as research and development. This

study develops an instrument for assessing mind map products to assess students' creative thinking skills, especially the senior high school level in biology subjects. The development model used in this research is ADDIE (Analyze, Design, Development, Implementation, and Evaluation). Random purposive sampling technique was used to choose the subjects for trying the instrument. The samples are three classes of XI grade science students of *Madrasah Aliyah Negeri* (MAN) 3 Sleman, Yogyakarta.

The data collected are in the form of quantitative and qualitative data. Quantitative data were collected from expert validation questionnaire for content validity and data from the result of students' mind map for validating the construct validity. Qualitative data were collected from expert recommendation during content validity.

The analysis (analyze) stage is a stage where researchers conduct a need assessment of products, by conducting interviews with biology teachers as well as observing the grids and daily test questions as well as the final semester assessment. Need assessment results on assessment instruments for mind map products to assess students' creative thinking skills at high school level in biology, namely instruments needed by biology teachers as a more practical and effective assessment tool in addition to project appraisal and practicum to assess thinking skills creative high school students in biology subjects. The results of the elaboration of aspects based on conceptual and operational definitions are then made as statement points, arranged in the instrument lattice. There are 21 statement items arranged in the initial design of this instrument. Scoring technique used is rating scale; 4 (very good), 3 (good), 2 (enough), and 1 (less). Each statement has its own assessment criteria, which are outlined in the assessment rubric.

The design phase is the stage where researchers design products in the form of assessment instruments of mind map product to assess students' creative thinking skills in biology subjects. The researchers designed the format of instrument grids, assessment rubrics, assessment sheets, and scoring techniques. Instrument design is formed based on conceptual and operational definitions. Conceptual and operational definitions also determine indicators that will be developed into statement items to measure students' creative thinking ability through mind maps (Table 1). Also, at this stage, researchers also plan who will become the instrument validator (rater), which will validate the contents of the instrument.

Table 1. Aspects, sub-aspects, and indicators of mind map to assess creative thinking skill

Aspect	Sub-aspect	Item Indicator
A. Originality	A.1 Making distinctive characteristics	A.1.1 Using keywords that are their own characteristics
		A.1.2 Drawing patterns that are distinctive
	A.2 Add a symbol combination	A.2.1 Add symbols or images
		A.2.2 Add a color combination
B. Fluency	B.1 Pour ideas	B.1.1 Write down all material concepts
		B.1.2 Write down problems in daily life
	B.2 Work fast	B.2.1 Pour the concept in a relatively short time
		B.2.2 Complete a mind map in a relatively short time
C. Flexibility	C.1 Develop the concept of material	C.1.1 Add personal opinion
		C.1.2 Add material references
	C.2 Add a solution to the problem	C.2.1 Write down solutions to problems in everyday life
D. Elaboration	D.1 Describe the concept in depth	D.1.1 Write in detail the concept of matter
	D.2 Describes concepts with relevant components	D.2.1 Write concepts with relevant components
	D.3 Contact line	D.3.1 The line can be read clearly
	D.4 Readability of idea relations between concepts	D.4.1 The relationship between ideas is easy to understand

Development is the next stage, in which the researchers start developing the instrument, as well as conducting content validation on the instrument. The method used to determine the content validity of mind-map product assessment instruments to assess the ability to think creatively is the Lawshe method. Rater or Subject Matter Expert who was involved in the determination of the content index of the content for this amounted to nine people. The formula used to determine the ratio of content validity ratio (CVR) is as follows:

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

Notes:
ne : number of rater which states essence
N : the total number of rater

The range of CVR values is -1.00 to 1.00 so that when CVR> 0.00, an item is declared valid because more than half (> 50%) rater or Subject Matter Expert states that the item is essential (Azwar, 2012, p. 114). Data regarding item eligibility through content validation, obtained by validation sheet. Instruments that have been validated in the content are revised, so they are ready to be tested in the field. Implementation of the instrument (implementation), is the next step. The instrument was tested in the field to assess mind maps by biology teachers. The instrument was tested to assess the mind map created by 95 remaining XI Science. Determination of the sample was done by simple random sampling technique. The results of the field trial then became the basis for construct validation of the instrument, as well as testing the reliability of the instrument.

Construct validity explains the extent of the consistency of test performance with the constructs in a particular theoretical consideration (Hudha & Mardapi, 2018). The test of construct validity and reliability is carried out at the evaluation stage, the results of which state the feasibility of the instrument used to assess mind maps to determine students' creative thinking ability in biology subjects. Construct validity is done by Exploratory Factor Analysis (EFA) and then confirmed by Confirmatory Factor Analysis (CFA).

Exploratory factor analysis or principal component analysis (PCA) is a factor analysis technique in which several factors will be formed in the form of latent variables that cannot be determined before the analysis is carried out (Yong & Pearce, 2013, p. 79). Exploratory factor analysis was carried out using the SPSS program.

This analysis begins by testing the adequacy of the sample used in the analysis through the KMO (Kaiser-Mayer-Olkin) test. If the KMO value is between 0.5 to 1, then it is indicated that factor analysis is appropriate to use and can be used for further analysis (Bilson, 2005, p. 123). Then, look at the anti-image value or see the MSA for each item. The MSA value for each variable can be seen diagonally on the anti-image correlation in the diagonal part of the matrix. If one or several initial variables individually have MSA values less than 0.5, the variable is excluded from the analysis process (Bilson, 2005, p. 123). Calculate the eigenvalue, which is then used to calculate the percentage of variance that is explained, while drawing the scree plot, and grouping the items on the factors formed after the analysis. Factors that have eigenvalues greater than or equal to 1 will be maintained, and factors that have eigenvalues less than 1 will not be included in the model because variables with values less than 1 are not better than the original variables (Supranto, 2004).

The reliability testing technique which was used is Cronbach's Alpha, which is a coefficient that describes how well the items in a set are positively correlated with each other (Azwar, 2004). Reliability coefficients (Cronbach's Alpha) range from a score of 0.00 to 1.00 with the assumption that the closer to 1.00, the more reliable the measuring instrument is (Azwar, 2004, p. 78).

Confirmatory Factor Analysis (CFA) is a method of factor analysis which is used when researchers know the structure of a latent factor (Laili & Otok, 2014). CFA was used to confirm that the estimates model that was formed during exploratory analysis was fit. CFA was done using Lisrel program. The result of the CFA produced a standardized loading factor (SLF) and was determined as

the construct validity. When the SLF value of the indicator is over 0.30, the indicator is then considered as significant (Igbaria, Zinatelli, Cragg, & Cavaye, 1997, p. 290). Making it simpler, the ADDIE-model steps which were undertaken by the researchers are presented in Figure 1.

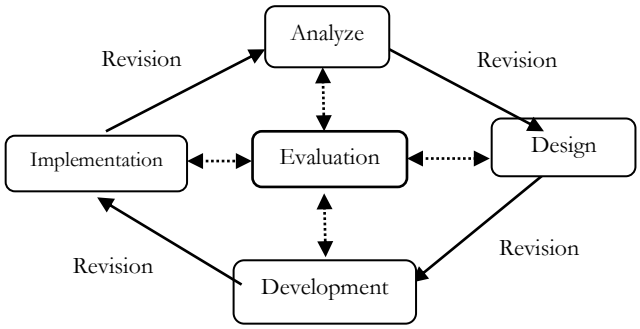


Figure 1. Steps for research and development of ADDIE models (Sugiyono, 2015, p. 39).

Findings and Discussion

Findings

The statement items that have been made subsequently are tested for content validity by the Lawshe method, involving nine raters. Calculations were done using the CVR formula. Of the 21 items submitted, there were three items declared invalid after the content validity was calculated, namely items 10, 14, and 16 (Table 2). The assessment results from rater are not just numbers but also some suggestions for improving grain quality. After revising, 18 items are valid to be tested.

The results of field trials or implementation of instruments in the field are used for the construct validity of the instrument. Construct validity was carried out using exploratory factor analysis. The analysis was carried out with SPSS program.

Table 2. Calculation results of content validity

Item	Rater									CVR	Validity
	1	2	3	4	5	6	7	8	9		
1	3	3	3	3	3	3	3	3	3	1	valid
2	2	3	3	3	3	3	3	3	3	0.78	valid
3	3	3	3	3	3	3	3	3	3	1	valid
4	3	3	3	3	3	3	3	3	3	1	valid
5	3	3	3	3	3	3	3	2	3	0.78	valid
6	3	3	3	3	3	3	3	3	3	1	valid
7	3	3	3	3	3	3	3	1	3	0.78	valid
8	3	3	3	2	3	3	3	3	3	0.78	valid
9	3	3	2	3	3	3	3	3	3	0.78	valid
10	2	3	2	3	2	2	3	3	2	-0.11	invalid
11	3	3	2	3	3	3	3	3	3	0.78	valid
12	3	3	3	3	3	3	3	3	3	1	valid
13	3	3	3	2	3	3	3	3	3	0.78	valid
14	2	2	3	2	3	3	3	1	2	-0.11	invalid
15	3	3	3	3	3	3	3	3	3	1	valid
16	2	3	3	3	2	2	3	1	2	-0.11	invalid
17	3	3	3	3	3	3	3	3	3	1	valid
18	3	3	3	3	3	3	3	3	3	1	valid
19	3	3	3	3	3	3	3	3	3	1	valid
20	3	3	3	3	3	3	3	3	3	1	valid
21	2	3	3	3	3	3	3	1	3	0.56	valid

Exploratory Factor Analysis

The first explanation begins with the results of KMO and Bartlett's test. The results of the sample adequacy factor analysis show chi-square values in Bartlett test of 456.795 with a degree of freedom 153 and a p-value of <0.01. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) was 0.839, which is greater than 0.5. If the KMO value is between 0.5 to 1, it means that the factor analysis is right.

MSA (Measure of Sampling Adequacy) can be seen from anti-image correlation diagonal matrices of each variable. The anti-image correlation value is marked by a superscript of 'a.' MSA analysis results show that all items have a relationship with other initial variables, so there no one item must be removed from the analysis process. It is in line with the statement of Bilson (2005, p. 123), if one or several initial variables individually have MSA values less than 0.5, then the variable is excluded from the analysis process. The method often used in exploratory factor analysis for factor formation is the principal component method with orthogonal rotation. The specific purpose of the principal component analysis method is to know the structure underlying the initial variables in the analysis and simplify the structure of the initial set of variables through data reduction.

Based on the eigenvalues and the components of the factor analysis variance with SPSS, it was found that the data obtained from the mind map product assessment using the instruments used contained 5 Eigenvalues greater than 1, so that it can be said that the instrument of mind map products contains 5 factors, but only 4 factors were expected. After two items that had the lowest MSA values were issued (item 4 and 11), re-analysis was carried out to produce the number of expected factors. The result of the re-analysis of KMO and Bartlett's Test is in Table 3.

Table 3. Result of KMO and Bartlett's Test

KMO and Bartlett's test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.857
Bartlett's Test of Sphericity	Approx. Chi-Square	456.795
	df	120
	Sig.	.000

Table 3 shows that the KMO value was 0.857; indicating that it is very good. Bartlett's Test Sphericity value was 456.795 at the degree of freedom 120 and signification 0.000, so it was very good. Based on eigenvalues, re-analysis produce four eigenvalues that were greater than 1, so the instrument of mind map product contains four factors as expected. Of the four factors, there was 56.116% variance explained.

Scree plot is a graph that shows the relationship between factors with their eigenvalues. Determining these criteria is done by plotting the eigenvalue of the number of factors to be extracted. The eigenvalues are plotted in the vertical direction, while the number of factors is plotted in the horizontal direction. The number of factors in this criterion is determined based on the decrease in the plot of the eigenvalues. On the scree plot generated from the analysis (Figure 2), shows four plot drop points, so that four factors were formed from 16 variables (items).

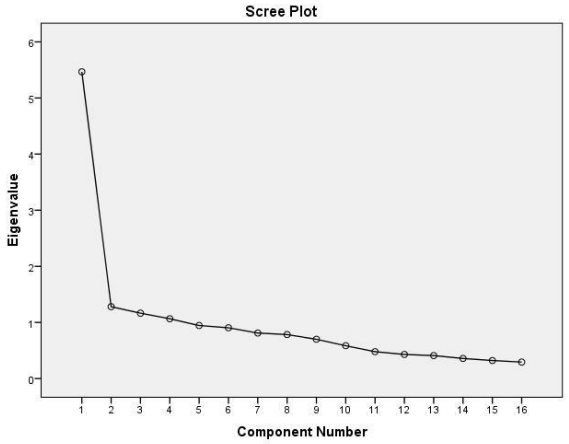


Figure 2. Scree plot

Four components or factors have been formed after the analysis process, then rotating these factors. The main objective of the rotation process is to achieve simplicity with factors and increase the ability of interpretation. The method used is the varimax method, which focuses its analysis on simplifying the column matrix factor. In this method, there is a tendency to produce several high factor loading values (close to -1 or +1) and some factor loading values close to 0 in each matrix column — a significant loading factor if ≥0.5.

We will look at the component matrices in Table 4 that have been rotated by variables or which items are worth ≥0.5, which will be incorporated into the factors or components that have been formed. Grouping items or variables in each component are presented in Table 5.

Table 4. Rotated component matrix[a]

	Component			
	1	2	3	4
item1	.061	.355	.703	-.239
item2	.737	.202	-.010	-.054
item3	.524	.108	.465	.162
Item5	-.607	-.234	-.125	-.152
Item6	.102	-.089	.827	.100
Item7	.328	.499	.113	-.240
Item8	.791	.082	.017	-.032
Item9	-.029	.759	.196	.224
Item10	.131	.092	.013	.818
item12	-.637	-.266	-.115	-.170
item13	-.551	-.275	-.374	-.080
item14	-.298	-.634	.003	-.067
item15	.185	.484	-.017	-.233
item16	.481	.585	.190	.142
item17	.387	.609	.025	.203
item18	.561	.379	.146	-.229

Notes:
Extraction method: Principal component analysis.
Rotation method: Varimax with Kaiser normalization.
a. Rotation converged in 5 literations.

Table 5. Grouping items

Factor	1	2	3	4
Item	2, 3, 8, 18	9, 16, 17	1, 6	10

Sixteen items were grouped into four factors formed during the analysis. Factor 1 consists of four items; factor 2 consists of three items; factor 3 consists of two items, and factor 4 consists of one item. There are six items, points 5, 7, 12, 13, 14, and 15, which do not belong to any group of factors. These six items will then be eliminated because there is no connection with the other items. Thus, from 16 items after factor analysis, six items must be removed.

The naming of each factor is the next step. Factor naming is based on the characteristics of each item or variable that is a member of the factor (Table 6). After we have the item which we will use, the next step is to find out the reliability of the instrument. Reliability of the instrument is 0.774.

Confirmatory Factor Analysis

The conceptual construct model design which had been formed during exploratory analysis was then analyzed with second-order CFA. The result is presented in Figure 3. The analysis result shows that the model designed complies with the goodness of fit statistics. The model fit of the instrument is indicated by the p-value = 0.68 (p>0.05), RMSEA = 0.00 (RMSEA<0.08) and Goodness of Fit Index = 0.94. Based on data, p-value and RMSEA were successfully met, so that it is indicated that this model was fit with the data.

Table 6. Naming factor

Items	Factor	Name of Factor
2. Adjust the proportion of the use of keywords and sentences	1	Flexibility
3. Drawing different pattern from others		
8. The concept of the material outlined is easy to understand		
18. Relationship between ideas help understand the whole contents of the mind map		
9. Mention examples of common problems	2	Elaboration
16. Components used in detailing concepts relevant to the material		
17. Write material with many branches		
1. Use keywords that are different from others	3	Originality
6. Use various colors in the mind map		
10. Explain the concept of the material in a maximum of 30 minutes	4	Fluency

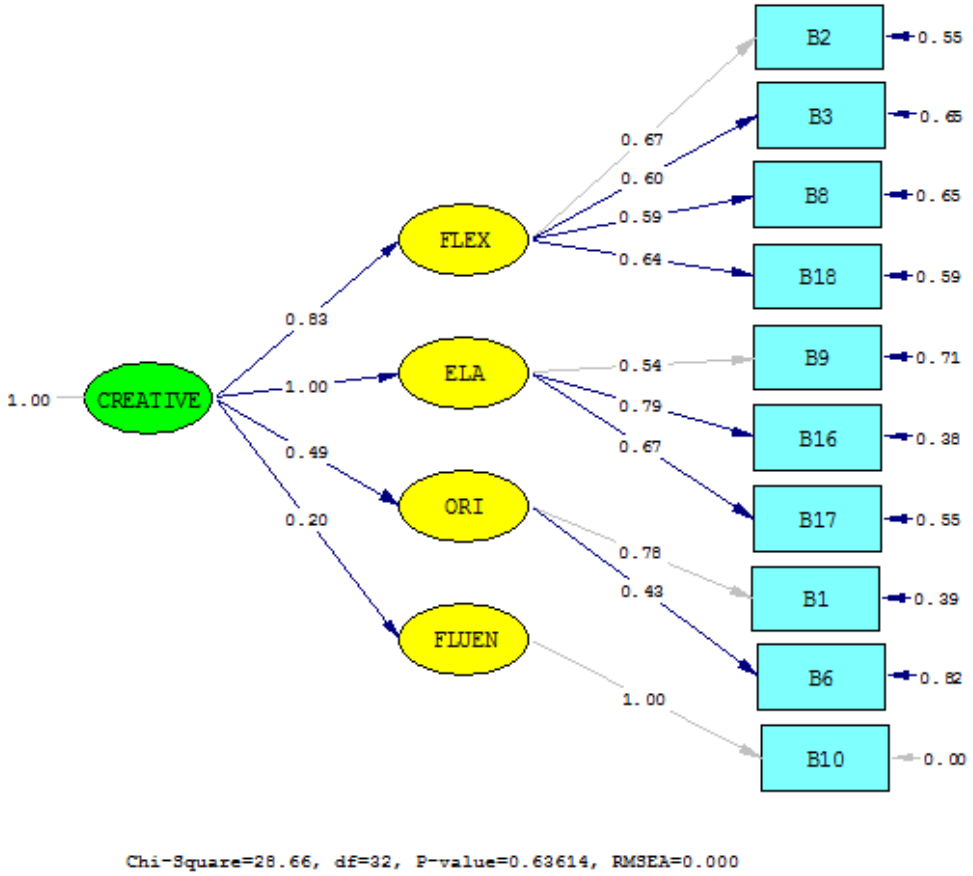


Figure 3. Path diagram of second order analysis output

Table 7. Result of second order CFA of the instrument of mind map product with 16 item

Aspects	Item	SLF	R^2	Notes
Flexibility (FLEX)	B2	0.67	0.61	reference item
	B3	0.60	0.35	item fit
	B8	0.59	0.35	item fit
	B18	0.64	0.41	item fit
Elaboration (ELA)	B9	0.54	0.29	reference item
	B16	0.79	0.62	item fit
	B17	0.67	0.45	item fit
Originality (ORI)	B1	0.78	0.39	reference item
	B6	0.43	0.18	item fit
Fluency (FLUEN)	B10	1.00	1	reference item

The result of the CFA produced a standardized loading factor and was determined as the construct validity. When the loading factor value of the indicator is over 0.30, the indicator is considered as significant (Igbaria et al., 1997, p. 290). Table 7 shows that all of the items have a loading factor value >0.30. Item 16 has the highest contribution to the measuring instrument with a loading factor of 0.79, while item 6 gives the smallest contribution with a loading factor of 0.43.

Discussion

All results of the analysis show that the sample size of 95 (3 of science classes) used in this factor analysis is sufficient. It can be seen from KMO value >0.5 (0.857). This is reinforced by the theory stated by Gable (1986) that the sample size or the number of respondents is 5 to 10 times the number of items. The item number or item in the trial is 18, so the minimum sample size is 90. In other words, a sample size of 95 is sufficient.

The approach used to determine the number of factors obtained in this study is based on eigenvalues, variance percentages, and scree plots. Eigenvalues show the number of variations related to a factor. Factors that have eigenvalues greater than or equal to 1 will be maintained and factors that have eigenvalues less than 1 will not be included in the model because variables with values less than 1 are not better than the original variables (Supranto, 2004). From the result of the eigenvalues, it is known that those variables in the instrument can explain 56.116% the variances. It means that the instrument has already been able to measure the aspect that will be measured.

Based on the findings, it is indicated that after construct validity was done, the items of the instrument were reduced, from 18 items to 10 items. The selected items are grouped into five factors. Every factor is named base on the same between items and aspect that will be measured.

Factor 1 is called flexibility because the variables incorporated in these factors consist of variables that show a person's ability to adjust keyword and picture that are in accordance with the students' own understanding to explain the material, and able to relate the whole material so the people who read it will understand. Factor 2 is called elaboration because the variables incorporated in these factors show how student able to detail a main idea of the material as detailed as possible so that all things about the material can be contained in the mind map and still relevance.

Factor 3 is called originality because it consists of variables which indicate the authenticity of the idea in making mind maps by using the keywords that are their own characteristic and using various colors which fit with their creativity. Factor 4 is called fluency, which is seen from using time in making a mind map.

From the findings, it is found out that to assess the ability to think creatively through mind map creation, we can see it by judging from the originality of ideas, fluency in pouring and connecting ideas and the speed of expressing ideas, how to elaborate a material, and flexibility in displaying ideas with their own approach. Those are same with previous research that creative thinking skill can be measured from originality, fluency, flexibility, and elaboration (Hartati et al., 2017; Kristiani et al., 2017; Putri, 2016).

Several items are not incorporated in the initial aspect or factor. This is the purpose of the construct validity with EFA. EFA is grouping the variables (items) that have the same correlation to measure the same construct (Hair, Black, Babin, & Anderson, 2010).

Half of the initial items in the instrument are not used after the construct validity test. Out of it all of the items have anti-image value >0.5. It can occur due to the lack of variants in the data collected. To construct validity test with EFA, every item has a minimal ideal value of factor loading, which is 0.55 for small scale test, and 0.35 for big scale test (Hair et al., 2010). The factor loading 0.55 if the trial subjects are 100 and 0.35 need 250 subjects. It is the shortcoming of the research. The research still needs more trial subject. Future research is suggested to use more subject trial to do the construct validity with EFA.

Reliability calculation of assessment instruments of mind-map product to measure students' creative thinking ability is done after conducting content validity and construct validity. The items that have been selected through the validity process are the estimated reliability estimates. Reliability calculations use the SPSS program with the Cronbach's Alpha technique. Cronbach's Alpha was chosen to estimate the instrument's reliability with a polytomous score (Retnawati, 2015, p. 91). The results of the reliability calculation of the mind-map product assessment instrument with ten items, using Cronbach's Alpha in the SPSS program amounted to 0.774. It shows that the reliability of the instrument is good or the instrument has been reliable.

Model Fit testing Goodness of Fit (GOF) is obtained to test whether the initial model has fitness with the sample data or not. The model fit of the instrument is indicated by the p-value = 0.68 (p>0.05), RMSEA=0.00 (RMSEA<0.08) and Goodness of Fit Index = 0.94. Chi-square value is 28.66 (χ^2=28.66) with a degree of freedom is 33 ($0 \leq \chi^2 \leq$ 2df).

Chi-square and p-value are to test the structural equation model fitness. Root Mean Square Error of Approximation (RSMEA), a measure of approximate fit of a model in the population, is related to differences in estimates (Riadi, 2018). Browne and Cudeck in Riadi (2018) classify 'close fit' if RSMEA value between 0.05 until 0.08 as 'adequate fit', and RSMEA between 0.08 until 0.10 as 'mediocre fit', but if RSMEA >0.10 could not accept. Hu and Bentler (1998) suggest that RSMEA value <0.06 as a cutoff criterion. Based on the result, RSMEA value for this instrument is in 'close fit' category.

The goodness of Fit Index (GFI) has a value between 0 until 1. GFI > 0.90 is 'good fit', 0.80 < GFI < 0.90 is 'marginal fit' or it can be accepted. Based on the result, GFI value is 0.94, so it means that it is in the 'good fit' category. Another evidence that the model is fit with the data is all of the items has loading factor value >0.3. Item 16 has the highest contribution to the measuring instrument with a loading factor of 0.79, while item 6 gives the smallest contribution with a loading factor of 0.43.

Item 16 is included in elaboration factor. Elaboration is an ability to develop an idea and add or detailing from an object, idea, or situation, so it makes it more interesting (Munandar, 2009). In other words, to know the creative thinking skill, it can be seen from how to elaborate an idea in a mind map.

The instrument developed can be one of way to assess student' creative thinking skill in biology subject, because that was a reliable instrument. This instrument also can be a reference for a teacher who wants to develop a product of assessment instrument.

Limitation in this research are as follows. (1) The use of classic tests compared to item response theory (IRT) due to the limited number of samples used does not allow the use of an IRT. (2) The use of EFA (exploratory factor analysis) and CFA (confirmatory factor analysis) is because the researchers formed new item indicators so that they did not know the latent variables where the indicators would group. CFA was also used to test whether the indicators that have been grouped based on the latent variable (construct) are consistent in the construct or not. (3) The try out of the instrument was not conducted in the school observed by the researchers when need analysis was done. The researchers tried out the instrument in MAN 3 Sleman, because of limited time. However, before the researchers do the tryout, they have interviewed the biology teacher in that school, who stated the same problems with the schools observed before.

Conclusion

From the findings and discussion, some conclusions are drawn. (1) Content validity results in the range number of CVR is 0.56—1.00, construct validity results to five factors that can assess students' creative thinking skill in biology subject, and the reliability estimate of the instrument is 0.774. Therefore, it is concluded that the instrument which is developed in this research is valid and reliable to assess students' creative thinking skill in biology subject. (2) The p-value = 0.68, RMSEA =0.00 and Goodness of Fit Index = 0.94. (3) This instrument can be tried out for future research with more subjects (100-250 subjects) to know how valid and reliable this instrument is in large scale subjects. (4) The shortcoming of this research is the instrument was tried out just in one school. It is better if it is tried out in more than one school. (5) This instrument can be a reference for a teacher who wants to develop a product of assessment instrument. The instrument produced in this research can be used by the teacher to assess students' creative thinking skill with the heterogeneous students as a formative assessment in biology subject. Because the instrument contains statements to measure the creative thinking skill in general, then this instrument can be used in all the material in biology subject.

Acknowledgment

The researchers would like to thank to Indonesia Endowment Fund for Education (LPDP) from the Ministry of Finance of Republic of Indonesia for always supporting the first author by scholarship.

References

Azwar, S. (2004). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.

Azwar, S. (2012). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.

Bilson, S. (2005). *Analisis multivariat pemasaran*. Jakarta: Gramedia Pustaka.

Buzan, T. (2011). *Buku pintar mind map*. Jakarta: PT Gramedia Pustaka Utama.

Çoban, S., & Tokatlı, E. S. (2017). The effect of mind mapping technique on students' achievements in music lesson and on their attitudes towards the mind mapping technique. *Education and Science*, *42*(190), 423–435. https://doi.org/10.15390/EB.2017.6856

DePorter, B., & Henarcki, M. (2010). *Quantum teaching*. Bandung: PT Mizan Pustaka.

Gable, R. K. (1986). *Development in the affective domain*. Boston, MA: Kluwer Nijhoff Publishing.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454. http://dx.doi.org/10.1037/h0063487

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hartati, D. R., Asrial, & Ernawati, D. W. (2017). Pengembangan instrumen penilaian berpikir kreatif peserta didik pada materi larutan asam dan basa kelas XI SMA/MA. *Artikel Ilmiah*, 1–10.

Herlina, L., & Qurbaniah, M. (2017). Analisis kemampuan berpikir kreatif siswa pada materi virus kelas X MAS Al-Mustaqim Sungai Raya 2. *Jurnal Bioeducation*, *2*(1), 11–14.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Hudha, S. A., & Mardapi, D. (2018). Developing an instrument for measuring the spiritual attitude of high school students. *REiD (Research and Evaluation in Education)*, *4*(1), 35–44. https://doi.org/10.21831/reid.v4i1.20304

Hyerle, D. (2012). Peta pemikiran (thinking maps) sebagai suatu bahasa transformasional untuk pembelajaran. In N. D. Hyerle & L. Alper (Eds.), *Peta pemikiran (thinking maps): Penelitian berbasis sekolah, hasil dan model untuk prestasi dengan menggunakan peralatan visual* (2nd ed.). (A. Cahyani, Trans.) Jakarta: PT. Indeks.

Igbaria, M., Zinatelli, N., Cragg, P., & Cavaye, A. L. M. (1997). Personal computing acceptance factors in small firms: A structural equation model. *MIS Quarterly*, *21*(3), 279–305. https://doi.org/10.2307/249498

Kristiani, K. D., Mayasari, T., & Kurniadi, E. (2017). Pengembangan ssesmen keterampilan berpikir kreatif siswa SMP Negeri 5 Madiun pada materi cahaya dan alat optik. In *Prosiding SNFA (Seminar Nasional Fisika dan Aplikasinya)* (pp. 102–108). Indonesia.

Laili, M., & Otok, B. W. (2014). Second-order confirmatory factor analysis pada kemiskinan di Kabupaten Jombang. *Jurnal Sains Dan Seni ITS*, *3*(2), 278–283. https://doi.org/10.12962/J23373520.V3I2.8161

Moore, K. D. (2015). *Effective instructional strategies: From theory to practice* (4th ed.). Los Angeles, CA: SAGE Publications.

Munandar, U. (2009). *Pengembangan kreatifitas anak berbakat*. Jakarta: Rineka Cipta.

Murley, D. (2007). Technology for everyone...: Mind mapping complex information. *Law Journal Library*, *99*(1), 175–183.

Putri, A. R., Paidi, P., & Subali, B. (2016). Kemampuan berpikir divergen keterampilan proses sains aspek biologi siswa SD berdasarkan lokasi sekolah. *Jurnal Pendidikan Biologi*, *5*(5), 39–50.

Putri, L. O. L. (2016). Mind map sebagai model pembelajaran menilai penguasaan

konsep dan alat evaluasi menilai kemampuan berpikir kreatif siswa. In *Proceeding of Seminar Nasional Pendidikan dan Saintek* (pp. 629–634). Surakarta: Universitas Muhammadiyah Surakarta. Retrieved from https://publikasiilmiah.ums.ac.id/handle/11617/7980

Regulation of Minister of Education and Culture, No. 14 of 2018, on Student Admissions (2018). Republic of Indonesia.

Retnawati, H. (2015). *Validitas reliabilitas dan karakteristik butir (Panduan untuk peneliti, mahasiswa, dan psikometrian)*. Yogyakarta: Parama Publishing.

Riadi, E. (2018). *Statistik SEM structural equation modeling dengan Lisrel*. Yogyakarta: Andi.

Santrock, J. W. (2014). *Psikologi pendidikan*. (D. Angelica, Trans.). Jakarta: Salemba Humanika.

Sari, A. A., & Jarnawi, A. D. (2008). Pengaruh pemberian tugas creative mind map setelah pembelajaran terhadap kemampuan kreativitas dan koneksi matematik siswa. In *Proceeding of Seminar Nasional Matematika dan Pendidikan Matematika*. Yogyakarta: Universitas Negeri Yogyakarta.

Sugiyono. (2015). *Metode penelitian & pengembangan*. Bandung: Alfabeta.

Supardi, Haryanto, & Suhendri, H. (2014). Pengembangan instrumen pengukuran karakter kreativitas siswa. In *Proceeding of Seminar Nasional Hasil-Hasil Penelitian (SNHP-IV) Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas PGRI Semarang* (pp. 290–297). Semarang.

Supranto. (2004). *Analisis multivariat arti dan interpretasi*. Jakarta: Rineka Cipta.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

Yusnaeni, Corebima, A. D., Susilo, H., & Zubaidah, S. (2017). Creative thinking of low academic student undergoing search solve create and share learning integrated with metacognitive strategy. *International Journal of Instruction*, *10*(2), 245–262. https://doi.org/10.12973/iji.2017.10216a

REiD (Research and Evaluation in Education), 5(1), 2019, 54-60
Available online at: http://journal.uny.ac.id/index.php/reid

Testing the parent-child communication program: Its effectiveness on developing children's social competences

[*1]Agus Mulyanto; [2]Alifah Indalika Mulyadi Razak
[1,2]Department of Early Childhood Education, Universitas Islam Nusantara
Jl. Soekarno-Hatta No. 530, Buahbatu, Kota Bandung, West Java 40286, Indonesia
[*]Corresponding Author. E-mail: agusmulyantouin@gmail.com

Submitted: 02 August 2018 | Revised: 16 April 2019 | Accepted: 07 May 2019

Abstract

Social competence has a central role in the development of early childhood. One of the central roles of children's social competence is its influence on children's academic social abilities, such as the ability to (1) receive learning information, (2) follow the school's rules, and (3) face increasingly complex academic challenges along with increasing school levels. This study aims to test the model of parent-child communication (PCC) in developing children's social competence. The exclusive communication model was tested to 250 children aged 4-6 years in West Java Province, which is divided into five research zones, West Java Zone 1 to 5. This study used an experimental design pre-test and post-test to determine the effectiveness of the parent-child communication program that was tested through observation and interview techniques consisting of 68 items of social competence. The results show that the PCC Program can effectively be applied by the collaboration of parents between fathers and mothers to optimize children's social competencies. The PCC program, which was not attended by both parents, would not be effective, for example, as happened in the West Java Zone 1, indicating the ineffectiveness of the PCC program because of the characteristics of parental activities that both work and do not have time to communicate with children. While in the other four West Java Zones PCC can be effective, because working parents want to take the time to interact and communicate actively with their children.

Keywords: *parent-child communication, social competences, children*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.20679

Introduction

Social competence is considered to play an essential role in child development. Izard et al. (2001) explain that social competence can affect children's academic abilities. Social competence includes the ability to know yourself, know the environment, manage interaction with the environment, share ability, gain prosocial skills, manage the challenges of learning in school, recognize the various emotional expressions, and make friends. The results of another study indicate that social competence can have a significant effect on the academic social life of a child (Kail, 2012). A child who has good social competence is predicted to have good performance at school (Papalia, Feldman, & Martorell, 2012). The result of a survey conducted to the teachers of the Pembina State Kindergarten in West Java shows that 4-to-6-year-old children who were in kindergarten groups still need maximum stimulation in order to improve their social competence (Mulyanto, Muchtar, Hanafiah, Hoerudin, & Razak, 2017), such as in the subcompetence of prosocial behavior, self-concept, emotional well-being, and academic-social life. In addition, Mulyanto et al. (2017) also argue that 50% of kindergarten children attending school at Pembina State Kindergartens in West Java Province have not shown maximum social competence.

It is clear that children's social competence is influenced by various community

This is an open access article under the CC-BY-SA license.

contexts (Kärtner, Keller, & Chaudhary, 2010). Various efforts are made to stimulate the development of children's social competence, one of which is by giving rewards in the form of praise, star symbol, or toy that can effectively improve children's prosocial behavior, which is, one of the indicators of social competences (Fabes, Fultz, Eisenberg, May-Plumlee, & Christopher, 1989). Besides, storytelling activities (Mincic, 2007), video games (Greitemeyer & Osswald, 2010), and lego games (Pang, 2010) are also proven to be an alternative to improve children's social competence. This form of stimulation is inseparable from the participation of parents and teachers to direct the children in developing their social competence, as evidenced by the research conducted by Berns (1997) that presents the involvement of teachers and parents as one of the successes of school programs. The problem that then arises is if parents and teachers have established good communication, then, how the communication between parents and children are performed, so that learning information to improve children's social competence in school can also be aligned with routine activities at home.

Answering the above questions, Jhon Bowlby's attachment theory introduces Parent-Child Communication (PCC) is a form of parent and child communication (Papalia et al., 2012; Schneider, Atkinson, & Tardif, 2001). The established communication involves an emotional bond between parent and children. It is hoped that there will be a mutual attachment contributed in establishing a high-quality relationship in accordance with the understanding of the concept of the attachment (Bus, Belsky, van Ijzendoom, & Crnic, 1997; Schneider et al., 2001). Through the PCC program, children feel safe and comfortable to tell their parents about what they feel and what they want to convey. Besides, parents can also ask children to do or behave in a certain way so that both parties can easily negotiate and find solutions to the events at hand. The PCC has been implemented in 2006 (Niles, Reynolds, & Nagasawa, 2006) as a form of program to convey cognitive information from parents to children, by using learning media as the intermediaries of parent-child communication program (Berns, 1997). Based on the previous experiences in implementing PCC, it is assumed that this program can also be effective in improving children's social competence.

Method

This study used a quantitative experimental design with the employment of before-and-after experiment design, or commonly known as pre-posttest design (Kumar, 2011). The population of this study consisted of 4-6-year-old kindergarten children who attend TK Negeri Pembina in West Java. These schools were selected to become the population because the characteristics of the environmental conditions and facilities are considered homogeneous since each TK Negeri Pembina has the same service and facility standards. The total population of the TK Pembina in the province was 50 schools, and the number of the kindergarten students was ranging from 80 to 100. Thus, the total student population was approximately 5,000 children. Therefore, the population needs to be limited through the determination of research samples by using random sampling (Kumar, 2011), so that the total study sample consists of 250 children from five zones in West Java province.

The participants of this study were 250 children from kindergarten in West Java province with characteristics of 4 to 6 years old, living with one father and one mother, and interacting with their father and mother every day. The research subjects were then grouped into 5 zones so that each zone consisted of 50 children. Consideration of the distribution of research zones is based on the geographical conditions of data collection in the province of West Java, including: (1) Zone 1 consists of kindergartens in Bandung Raya; (2) Zone 2 consists of TK in the northern part of West Java; (3) Zone 3 consists of TK in the southern part of West Java; (4) Zone 4 consists of eastern part of West Java; (5) Zone 5 consists of the western part of West Java.

This study aims to examine the effectiveness of parent-child communication programs to improve children's social competence. The PCC program is applied on the

questions under the following cases: (1) communication when the child goes to school; (2) communication when the child comes home from school; (3) communication between day and evening; (4) communication when children have their dinner; and (5) communication when the child is at bedtime. The five-forms parent-child communication program is carried out by parents consisting of father and mother for ten days. The research can be said to be effective if the PCC program can improve children's social competence, which consists of 68 items measured through observation and interview. Observation applied to items that can be observed, such as the ability to play with friends, the ability to help friends, and other items. Interview measurement instruments applied to items that cannot be observed, such as the ability to distinguish sad, happy, disgusting, fearful, and other facial expressions. The technique of collecting data using observations and interviews was carried out by the class teacher of each child during the pre-test and post test process. Data analysis was carried out by agreeing to the parent-child communication program through the following hypothesis testing:

H_0: There is an influence of parent-child communication intervention on child social competency

H_1: There is no influence from parent-child communication intervention on child social competency

The data were collected using observation and interview techniques consisting of 68 items of valid child social competency indicators collected in the pre-test, which were then intervened at the treatment stage, and tested for the effectiveness at the posttest, with the following research steps: (1) Implementation of social competence pre-test for children aged 4-6 years old; (2) Implementation of parent-child communication model intervention; and (3) Post-test social competence of children aged 4-6 years. Each step is elaborated as follows.

Implementation of Social Competence Pre-Test for 4-to-6-Year-Old Children

This activity examines children's social competence, which consists of 68 items that have been tested for validity and reliability using expert judgment. The implementation of the pre-test of social competence employs observation and interview instruments. Children are conditioned in playing activities and daily routines at school. In playing the activities, the indicators that can be observed are, for example, indicators number 1, 2, 3, 5, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, and 23. Each of them measures the following ability: the ability to play with friends, resolve conflicts with friends, ask about feelings of friends, like playing with friends rather than playing alone, help friends, behave well with friends, work with friends, share things, share ideas, resolve conflicts, show concern for younger children, throw jokes, distinguish between right and wrong behaviors, show sympathy, show empathy, play with older friends, lead friends, and be led by friends (Kail, 2012).

Meanwhile, the ability indicators that can be raised through the interview process are shown through the items number 4, 7, 10, 14, 18, 28, 29, 30, 31, and 35, each of which measures the ability to tell the teacher about friends' feelings, know the signs of friends who need help, show attention to friends, give advice to persuade friends, have the initiative to create a game or activity, choose the role or task given by the teacher, express choices, tell others about something, plan an action, and know the purpose of a rule (Mulyanto et al., 2017).

Implementation of Parent-Child Communication Model Intervention

The treatment is carried out simultaneously to five zones in the province of West Java. The research team asked the principal and teachers to socialize the parent-child communication (PCC) program to the parents of students in kindergarten in West Java province. This treatment took place for ten days and was followed by the parents of students (Knafo & Plomin, 2006) consistently in the form of activities (1) communication when the children go to school; (2) communication when children go home from school; (3) communication between day and evening; (4) communication when children have dinner;

and (5) communication when the the children are going to sleep. The implementation of this intervention was carried out by children from families of one father and one mother and consistently divided role assignments in this intervention activity.

Post-Test Social Competence of Children Aged 4-6 Years.

This activity examines children's social competence consisting of 68 items by observing and interviewing children in play activities and daily routines at school, after being given a parent-child communication program intervention. The items of social competency indicators tested were the same as the indicator items in the pre-test activities, such as the indicator items number 24, 25, 26, 27, 32, 33, 34, 42, 43, 45, 46, 50, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 68, each of which measures the ability of being enthusiastic in school, following the rules, giving focused attention for 2 minutes, giving focused attention for 8 minutes, being brave to be left by the family members when the students are at school, running routine activities, completing two to three assignments at school, maintaining their own property, taking care of other people's goods, protecting themselves from strangers, showing the things they need, preserving the environment, controling anger, controling anxiety, controling the fear of scary things, controling the fear of darkness, controling the fear of certain situations, controling pain, controling sadness, controling the shyness to be able to have monologue, playing games, and showing confidence.

Meanwhile, the indicators of ability that can be measured through the interview process is on items number 36, 37, 38, 39, 40, 41, 44, 47, 48, 49, 51, 52, 53, 54, 55, 56, and 67, which respectively measure the ability to distinguish men and women based on their role and physical appearance, have a tendency towards favorite objects, make positive calls to his friends, mention things they like and dislike to be able to distinguish goods by themselves and those of others, distinguish comfort and discomfort with their physical care, distinguish comfort and discomfort in their physical appearance, follow the rules for their health, recognize the happy, sad, angry, fearful, surprised, disgust, and proud facial expressions (Csoti, 2009).

Findings and Discussion

Early childhood social competence is a form of children's school readiness (Setiawati, Izzaty, & Triyanto, 2017). Furthermore, the research in 2017 explained that children's social-emotional abilities influence almost all the development of the children. School readiness is something that can be intervened through collaboration between parents and teachers. The parent-child communication program is one of the interventions that can be applied to improve children's social competence, which has a long-term influence on school readiness.

The parent-child communication program applied in this study involves fathers and mothers who have children aged 4-6 years. The involvement of parents in implementing this program requires good cooperation. This program is carried out for ten days through five forms of communication activities. The problem is the research subject has a different condition of father and mother activities. Some children have fathers and mothers who both work from morning to night; there is also the condition of the father working from morning to night, and the mother who stays at home as a housewife. Then some children also have fathers who work and go home once a month. There are also children with conditions where their father works and returns once a month, and the mother also works from morning to night. Therefore, the implementation of the parent-vertical communication program results is very diverse, spread across five zones of West Java province.

Characteristics of research subjects in West Java 1 zone (Table 1) came from kindergarten throughout Bandung Raya, most of the parents of students had activities working outside the home. The PCC program is carried out for ten days and must involve both parents. In West Java Zone 1, out of 50 parents, only 15 parents were willing to consistently join the PCC program, ranging from communication before leaving for school, communication at home from school, communica-

tion between the afternoon and evening, communication at dinner, and communication before bedtime which were carried out completely by father and mother. The results show that the PCC program was less effective in the West Java Zone 1, which was indicated by 29 children experiencing an increase in social competence (mean positive ranks 29b = 26.81) with a total value of 777.50, 19 children experiencing a decrease in grades (mean negative ranks 19a = 20.97) with a total value of 398.50, and 2 others have the same pre-test and post-test scores (ties = 2c). Based on the results of the pre-test and post-test, it was concluded that H_0 was accepted (Asymp.Sig (2-tailed) = 0.052> α = 0.05) so that there was no effect of the PCC program on the average social competence of early childhood.

The characteristics of the research subjects in West Java Zone 2 (Table 1) consisted of kindergarten in the northern region of West Java province, namely Subang, Purwakarta, and Karawang regions, amounting to 50 children. Participation in the research subject in this zone was considered very active because most of the children were accompanied by their mother at home (only the father works) so that the communication runs effectively. It is evidenced by the results of research showing that 35 people experienced an increase in social competence (mean positive ranks 35b = 26.24) with a total value of 918.5. Meanwhile, 11 children experienced reduced scores (mean negative ranks 11a = 14.77) with a total value of 162.5, and four others had the same pre-test and post-test scores (ties = 4c). Based on the results of the pre-test and post-test, it was indicated that H_0 was rejected (Asymp.Sig (2-tailed) = 0.000 <α = 0.05), so that there was an effect of the parent-child communication program on the average social competence of early childhood.

The effectiveness of the parent-child communication program is also tested in the southern West Java Zone 3 (Table 2), namely kindergarten in Garut, Cianjur, and Sukabumi regions. Even though 20 parents stated that they worked with their father and mother, they were willing to set the time and duties in joining this PCC program. Thus, 50 children were intervened effectively by their parents for ten days through the PCC program. This can be seen from the results of a study that showed 32 children experienced an increase in social competence (mean positive ranks 32b = 30.09) with a total value of 963. Meanwhile, 18 children experienced a decrease in value (mean negative ranks 18a = 17.33) with a total value of 312 while no single sample had the same pre-test and post-test scores (ties = 0c). Based on the results of the pre-test and post-test, it was indicated that H_0 was rejected (Asymp.Sig (2-tailed) = 0.002 <α = 0.05), so that there was an effect of the parent-child communication program on the average of early childhood social competence.

Table 1. Result of PCC program in West Java for Zone 1 and 2

	West Java 1			West Java 2		
	N	**Mean Rank**	**Sum of Ranks**	**N**	**Mean Rank**	**Sum of Ranks**
Negative ranks	19[a]	20.97	398.50	11[a]	14.77	162.50
Positive ranks	29[b]	26.81	777.50	35[b]	26.24	918.50
Ties	2[c]			4[c]		
Total	50			50		

Table 2. Result of PCC Program in West Java for Zone 3 and 4

	West Java 3			West Java 4		
	N	**Mean Rank**	**Sum of Ranks**	**N**	**Mean Rank**	**Sum of Ranks**
Negative ranks	18[a]	17.33	312.00	19[a]	16.92	321.50
Positive ranks	32[b]	30.09	963.00	28[b]	28.80	806.50
Ties	0[c]			3[c]		
Total	50			50		

Improvement on children's social competence also occurs in the eastern West Java Zone 4 (Table 2), namely Ciamis and Sumedang areas, indicated by an increase in scores of 28 children (mean positive ranks 28b = 28.80) with a total value of 806.5. Meanwhile, 19 children experienced a decrease in value (mean negative ranks 19a = 16.92) with a total score of 321.5, while three samples had the same pre-test and post-test scores (ties = 3c). Based on the results of the pre-test and post-test, it was indicated that H_0 was rejected (Asymp.Sig (2-tailed) = 0.01 <α = 0.05), so that there was an effect of the PCC program on children's social competence.

Table 3. Result of PCC program in West Java Zone 5

	West Java 5		
	N	**Mean Rank**	**Sum of Ranks**
Negative ranks	19[a]	18.11	344.00
Positive ranks	30[b]	29.37	881.00
Ties	1[c]		
Total	50		

Good communication skills between parents and children are also marked in the West Java Zone 5 (Table 3), namely Bogor and Depok regions of 50 children experiencing an increase in social competence by 30 people (mean positive ranks 30b = 29.37) with a total value of 881. Meanwhile, 19 children experienced a decrease in value (mean negative ranks 19a = 18.11) with a total value of 344, while one sample had the same pre-test and post-test scores (ties = 1c). Based on the results of the pre-test and post-test, it was indicated that H_0 was rejected (Asymp.Sig (2-tailed) = 0.008 <α = 0.05), so that there was an effect of the PCC model on the average social competence of children.

Based on the five data compiled from five areas in the province of West Java and involving 250 children and 250 pairs of parents, it is proven that the parent-child communication model consists of five forms of communication, namely: (1) communication when children go to school; (2) communication when children go home from school; (3) communication between day and evening; (4) communication when children have a dinner; and (5) communication when the child is going to sleep. This model can be proven effective in improving early childhood social competence.

Conclusion

This study aims to examine the effectiveness of parent-child communication programs on the development of children's social competence. Children's social competence has a central position in their growth and becomes one of the main factors in children's readiness, so it requires good collaboration between parents and teachers to make it developed well. Through this parent-child communication program, parents who have diverse activities are asked to maximize their communication with children, especially in terms of improving children's social competency, which consists of 68 behavioral items.

The research findings show that the parent-child communication program can effectively improve children's social competencies consisting of 68 items. The PCC program can be implemented in the form of communication carried out by parents and children consisting of (1) communication before going to school; (2) communication when going home from school; (3) communication between the afternoon and evening; (4) communication at dinner; and (5) communication before going to bed. The five parent-child communication models are very effective in the eastern, western, southern, and also northern province of West Java as well as the condition of the research subjects with many moms and dads who spend their time at home.

Acknowledgment

The researchers deliver their gratitude to the Ministry of Research and Higher Education of the Republic of Indonesia for the research grant program so that the research results can be published.

References

Berns, R. M. (1997). *Child, family, school, community: Socialization and support* (4th ed.). Belmont, CA: Holt, Rinehart, and Winston, Inc.

Bus, A. G., Belsky, J., van Ijzendoom, M. H., & Crnic, K. (1997). Attachment and bookreading patterns: A study of mothers, fathers, and their toddlers. *Early Childhood Research Quarterly*, *12*(1), 81–98. https://doi.org/10.1016/S0885-2006(97)90044-2

Csoti, M. (2009). *Developing children's social, emotional, and behavioural skills*. New York, NY: Continuum International Publishing.

Fabes, R. A., Fultz, J., Eisenberg, N., May-Plumlee, T., & Christopher, F. S. (1989). Effects of rewards on children's prosocial motivation: A socialization study. *Developmental Psychology*, *25*(4), 509–515. https://doi.org/10.1037/0012-1649.25.4.509

Greitemeyer, T., & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology*, *98*(2), 211–221. https://doi.org/10.1037/a0016997

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *12*(1), 18–23. https://doi.org/10.1111/1467-9280.00304

Kail, R. V. (2012). *Children and their development* (6th ed.). Upper Saddle River, NJ: Pearson Education.

Kärtner, J., Keller, H., & Chaudhary, N. (2010). Cognitive and social influences on early prosocial behavior in two sociocultural contexts. *Developmental Psychology*, *46*(4), 905–914. https://doi.org/10.1037/a0019718

Knafo, A., & Plomin, R. (2006). Parental discipline and affection and children's prosocial behavior: Genetic and environmental links. *Journal of Personality and Social Psychology*, *90*(1), 147–164. https://doi.org/10.1037/0022-3514.90.1.147

Kumar, R. (2011). *Research methodology: A step-by-step guide for beginners*. London: SAGE Publication.

Mincic, M. S. (2007). *Dialogic reading with emotion-laden storybooks: Intervention methods to enhance children's emergent literacy and social-emotional skills*. Doctoral Dissertation. Department of Psychology, George Mason University.

Mulyanto, A., Muchtar, H. S., Hanafiah, Hoerudin, C. W., & Razak, A. I. M. (2017). Children's social competences: An early detection. *International Journal of Management and Applied Science*, *3*(12), 70–73.

Niles, M. D., Reynolds, A. J., & Nagasawa, M. K. (2006). Does early childhood intervention affect the social and emotional development of participants? *Early Childhood Research and Practice*, *8*(1).

Pang, Y. (2010). Lego games help young children with autism develop social skills. *International Journal of Education*, *2*(2), 1–9. https://doi.org/10.5296/ije.v2i2.538

Papalia, D. E., Feldman, R. D., & Martorell, G. (2012). *Experience human development* (12th ed.). New York, NY: McGraw-Hill.

Schneider, B. H., Atkinson, L., & Tardif, C. (2001). Child-parent attachment and children's peer relations: A quantitative review. *Developmental Psychology*, *37*(1), 86–100.

Setiawati, F. A., Izzaty, R. E., & 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. https://doi.org/10.21831/reid.v3i1.13663

REiD (Research and Evaluation in Education), 5(1), 2019, 61-74
Available online at: http://journal.uny.ac.id/index.php/reid

An analysis of Javanese language test characteristic using the Rasch model in R program

[*1]Muchlisin; [2]Djemari Mardapi; [3]Farida Agus Setiawati
[1,2,3]Department of Educational Research and Evaluation,
Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Karangmalang, Depok, Sleman, Yogyakarta 55281, Indonesia
[*]Corresponding Author. E-mail: muchlisinjanuary@gmail.com

Submitted: 28 February 2019 | Revised: 02 May 2019 | Accepted: 07 May 2019

Abstract

One skill required to solve a problem in the 21st century is communication. Two international languages that are important in communication and thought at school are English and German language. However, beside international language, the local language, such as the Javanese language, is also essential and need to be maintained. The purpose of this study is to analyze the Javanese language test characteristics. This study was explorative research with secondary data collected by documentation of 220 students responses to the 50 multiple choice item of Javanese language test in the 11th grade of vocational high school. Data were analyzed using the Rasch model assisted by R program. Rasch model fits the data with 42 items after three times calibration. Based on difficulty level, ICC, and items reliability, there were 28 of 42 items (66.67%) that were good. This study finds out that generally, the Javanese language test is in the moderate category of difficulty. Hence, the need of evaluating the Javanese language test to make a better test that gives more accurate information about examinees' ability is crucial. The evaluation of the Javanese language test can be used to plan the next learning to get better Javanese language learning.

Keywords: *Javanese language test, Rasch model, R program*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.23773

Introduction

In the 21st century, there are some skills that are required. One of these skills is communication (Dede, 2010, pp. 7–8; Trilling & Fadel, 2009, p. 54; Zubaidah, 2017, p. 1). We need language to carry out communication. Some international languages are important, taught in the school, and widely used in the world, such as English, German language, Chinese language, etc. Beside international language, the local language, such as the Javanese language, is important and need to be maintained.

Central Java and Yogyakarta Special Region, two provinces in Indonesia, are very rich in terms of tradition and culture of Java. One of these traditions is the Javanese language that is used to speak to each other in daily life. This is why the Javanese language lesson at school, especially in Java, still be held nowadays. At every end of the semester, a test is conducted to assess students ability in the Javanese language.

The assessment of the Javanese language test can be carried out by analyzing test characteristics, which was begun by collecting the information about the previous results of the test score (Sumintono & Widhiarso, 2015, p. 12). Besides to give a score to the students, the students' response can also be used to predict or explain the students' ability and item characteristic by analyzing test characteristic based on the Item Response Theory (IRT).

Test is very important both for teacher and students. A test can be used to classify the weakness in terms of verbal skills, me-

This is an open access article under the CC-BY-SA license.

chanical skills, etc. (Allen & Yen, 1979, p. 1). Besides, a test is a powerful method of data collection with an impressive array for gathering numerical data rather than verbal kind (Cohen, Manion, & Morrison, 2007, p. 414). A test is defined as the standardized procedure for sampling behavior and describing it with categories or scores (Gruijter & van der Kamp, 2008, p. 2). The essential features of a test are a standardized procedure, a focused behavioral sample, and description in term of scores or categories mapping (Gruijter & van der Kamp, 2008, p. 2). The result of the test (scores) can be used to predict or explain the item and test performances (Lord & Novick, 2008, p. 358). Thus, the Javanese language test has to be analyzed in terms of its characteristics to get a better test in the next chance that can reach the test goal and give more accurate information about the examinee's ability.

The test has some uses. Five uses of a test include classification, diagnosis and treatment planning, self-knowledge, program evaluation, and research (Gregory, 2015, p. 29). A test can be a useful tool, but it can also be dangerous if misused (Allen & Yen, 1979, p. 5), depending on our professionality in ensuring the use of the test accurately and as fairly as possible. Many extraneous factors can influence the test (Gregory, 2015, p. 31). Several sources that may influence the test are the manner of administration, the test characteristic, the testing context, examinee's motivation and experience, and the scoring method (Gregory, 2015, p. 31).

In a test, some plannings need to be prepared, including identifying the purposes, the test specifications, and selection of the contents, considering the form, the writing test, the layout, the timing, and planning the scoring of the test (Cohen et al., 2007, p. 418). We can make a good Javanese language test by paying attention to the planning and some influencing factors. Besides, a good result of the test, which is accurate, rich, and beneficial for evaluation will be obtained by analyzing the characteristics of the items or test of Javanese language using Item Response Theory (IRT).

There are some alternative ways to analyze test characteristics, including classical test theory (CTT) and item response theory (IRT). In CTT, it is difficult to analyze a test with a large amount of calculation to get useful information (Baker, 2001, p. 1). Besides, CTT has some weakness, such as the result of the measurement depends on the test characteristic used, item parameter depends on the examinee's ability, and the error measurement provided is limited for group measurement instead of individual information (Mardapi, 2017, p. 187). In CTT, if test is 'hard', the examinee ability will below; it is 'easy', the examinee ability will be higher (Ronald K. Hambleton, Swaminathan, & Rogers, 1991, p. 2). Therefore, CTT is considered to be not effective to analyze the Javanese language test.

The weakness of CTT is that it can be covered by IRT. IRT is one of the modern psychometric theories that provide useful tools for ability testing (Harrison, Collins, & Müllensiefen, 2017, p. 1). IRT is a powerful tool used to solve a major problem of CTT (Downing, 2003, p. 739). Item response theory (IRT) models, including Rasch, show the relationship between the ability of test participants from latent trait (e.g., Javanese language skills) and the opportunity to master the given items (answer the items correctly) in the form of logistic models (Finch & French, 2015, p. 181). IRT has 3 assumptions (Finch & French, 2015, p. 181; Mardapi, 2017, p. 187). These are monotonicity, unidimensionality, and local independence.

CTT has served development well in a test over several decades, but IRT has become mainstream rapidly as the theoretical measurement basis (Embretson & Reise, 2000, p. 3). The feature of IRT is specification of a mathematical function relating probability of an examinee's response on a test item to an underlying ability (Embretson & Reise, 2000, p. 8; Finch & French, 2015, p. 177; Gruijter & van der Kamp, 2008, p. 133; R K Hambleton & Swaminathan, 1985, p. 9; Ostini & Nering, 2006, p. 2; Reckase, 2009, p. 68; van der Linden & Hambleton, 1996, p. iii). In other words, the function describes in probabilistic terms, a person with low and high ability give

a different response (Ostini & Nering, 2006, p. 2). IRT is an important thing that can solve the problem of dealing the relationship between ability (examinee's mental traits) and response (performance) to the item (Lord & Novick, 2008, p. 397). IRT is used in so many education fields, not only in social science, even in medical education, it has some potential benefits (Downing, 2003, p. 739). In the IRT, some information about the test characteristic can be gained accurately, so that analyzing the Javanese language test using IRT needs to be conducted.

One of the models in IRT is the Rasch model. The Rasch model was developed by Georg Rasch, a Danish mathematician, in 1960 (Hailaya, Alagumalai, & Ben, 2014, p. 301; Jambulingam, Schellhorn, & Sharma, 2016, p. 50; Mallinson, 2007, p. 1; Young, Levy, Martin, & Hay, 2009, p. 545). There are some points of view about the Rasch model. Rasch model is a special case of one-parameter logistic (1 PL) model with item discrimination value is set equal to 1 (Finch & French, 2015, p. 181). Discrimination shows the ability of an item to differentiate among examinees ability (Finch & French, 2015, p. 181). The Rasch model can be expressed as:

$$P(x_j = 1|\theta, b_j) = \frac{e^{(\theta - b_j)}}{1 + e^{(\theta - b_j)}} \qquad (1)$$

In equation (1), x_j is the response to the item j with 1 being correct in the context of an achievement test. θ represents an individual ability, and b_j is the difficulty level of item j.

Analysis of the Javanese language test using Rasch model has practical benefits. We can check the model fits the data. Rasch model can define the probability of a specified response in relation to examinee's ability and item difficulty of a Javanese language test (Hailaya et al., 2014, p. 301; Jambulingam et al., 2016, p. 50). Using Rasch model, there is no need to differentially weight items to produce a total score that gives the maximum possible amount of information about latent trait; the number-right score is the best possible total score to use (Allen & Yen, 1979, p. 260). Rasch model produces the latent-trait (Javanese ability) and the item difficulty scale that have desirable. Analyzing the Javanese language test using the Rasch model can be done by the R program.

The Javanese language test in the school has to be analyzed the characteristic using the Rasch model in IRT by R program to get some information. This information can gained from the Item Characteristic Curves (ICC). ICC can provide the probability of the examinees at a given ability level of answering each item correctly (Hambleton & Swaminathan, 1985, p. 13). Beside ICC, there are the other important information about the items or the test that we can get by using the Rasch model in IRT.The Javanese language test in the school has to be analyzed the characteristic using Rasch model in IRT by R program to get some information. This information can be collected from the Item Characteristic Curves (ICC). ICC can provide probability of the examinees at a given ability level of answering each item correctly (Hambleton & Swaminathan, 1985, p. 13). Beside ICC, there are the others important information about the items or the test that we can get by using the Rasch model in IRT.

There are many studies of IRT application. They compared the use of IRT and CTT or studied the application of IRT to analyze the test characteristic. A study conducted by Downing (2003) contrasts the IRT with CTT and explores the benefit of IRT application in typical medical education settings. Downing just compares these models and explore the benefit of IRT theoretically; he did not go further discussing the application of IRT in the analysis. In this study, IRT was used to analyze the test by the Rasch model in the R program. Essen, Idaka, and Metibemu (2017) analyze the model-data fit in IRT using Bilog and IRTPRO program. They used two programs to analyze the model-data fit, but in this study, one model in one program was used to analyze the model's fit data, item fit model, the difficulty level of the items, items characteristics curve (ICC), item information curve (IIC), test information curve (TIC), the information given by each item, and the Javanese ability distribution. More complex information would be revealed in this study.

The study of Purnama (2017) was conducted to understand the characteristics of Accounting Vocational Theory test items by IRT using BILOG Program. In this study will analyze the characteristics of the Javanese language test using the Rasch model in the R program. Purnama's study analyzes the test using 2 PL, employing the Rasch model, which is the special case of 1 PL. Purnama's study did not use the ICC to analyze the item characteristics, while in this study, ICC will be used. Another study conducted by Setiawati, Izzaty, and Hidayat (2018b, 2018a) using IRT to analyze the test employs Bilog program, while this study employs the R program. A study by Iskandar and Rizal (2018) has some relevancy with this study. These studies use a program to conduct analysis. In their study, they analyze the validity, reliability, difficulty level, and the other cases, but not the items and test characteristic curve, the information functions, the ability average of examinees, etc. Those aforementioned studies used CTT, while this study uses IRT. It is hoped that this study would present findings which can contribute to analyzing the characteristic of the Javanese language test, so that there would be an evaluation for the Javanese language test to get a better one.

The Javanese language test will be analyzed by IRT. Analyzing the Javanese language test will be more accurate and can be used to estimate the relationship between the examinee ability and the examinee response to the items of the Javanese language test. Analyzing the Javanese language test using IRT will produce the analysis not just for the overall test, but also for individual items characteristic. The characteristics of item and test (IIC and TCC) estimate how accurate the Javanese language test will give us the information (IIC and TIC) and the other characteristics. Based on the explanations, the researchers decided to analyze the Javanese language test characteristics based on item response theory using the Rasch model in the R program.

Method

This study is explorative research, that is research which aims at finding the fact and characteristics systematically and accurately about atheJavanese language test (Arikunto, 2010, p. 14). The characteristics of the Javanese language test were analyzed using the Rasch model in the R program. This research was conducted in Yogyakarta from May to June 2018.

The data analyzed in this study are secondary data. The data were collected by the documentation method, which is collecting the answer sheet of 220 students' responses to the Javanese language test in Depok 1 Vocational High School, Yogyakarta. The Javanese language test consists of 50 multiple choice items.

The instrument unit, the Javanese language test, was made by the Javanese language teacher. Then, the researchers summarize the responses in the dichotomy data table. The wrong responses are denoted by 0, and the true responses are denoted by 1. The item number 1 was symbolized with B1, item number 2 was B2, item number 3 was B3, and so on. The data of the Javanese language test were analyzed based on IRT using Rasch model in the R program.

After the data were collected and analyzed using the Rasch model in the R program, some findings are gained. It described how the characteristics of the Javanese language test told us the probability of an examinee's response on the test item to an underlying ability (Javanese language ability). The researchers analyzed the model fits of the overall data, the difficulty level, and item fits of the model, ICC, TCC, IIC, TIC, item information, the Javanese language ability distribution, and the descriptive statistics for the Javanese language ability.

The model fits the overall data. The goodness of fit model was conducted to test whether the Rasch model fits with the overall data, whereas item fits model was done to test whether the model fits for individual items as well. Both will be fit if the p-value more than 0.05. If the Goodness of Fit Model has not met the fit criteria, then the item fits model would be conducted, and the items that did not fit would be removed. Then, the goodness of fit of the remained items would be re-

analyzed until the criteria were met, and we can continue to the next analysis.

In practice, the researchers set the category, e.g., a difficult level is said to be good if it has a difficulty value ranging from - 2.0 to 2.0 (Hambleton & Swaminathan, 1985, p. 107). In this study, an item can be said a good item if have difficulty level from – 3.0 to 3.0. The ICC will show about how the relationship between examinee ability with the true response probability, whereas TCC shows the relationship between examinee ability and the true score (sum of the true response probability). The IIC and TIC show the information that we can get based on the item or test for certain examinee ability. The item information is useful for item selecting. The criteria of the reliable item are if the item information value more than 0.5. The Javanese language ability distribution and descriptive statistics are all about examinee ability in this test. All of the information would explore the Javanese language test characteristics in this study.

Findings and Discussion

After the data were collected and analyzed, some results are gained. It describes how the characteristics of the Javanese language test told us the probability of an examinee's response to the test item to an underlying ability (Javanese language ability). It can be seen from model fits data, the difficulty level, and item fits model, ICC, TCC, IIC, TIC, the distribution of Javanese language ability, etc.

The first step of the analysis of the characteristic of the Javanese language test is the assessment of the model fit for the Rasch model. We have to make sure that overall model fit for Rasch model. It can be said that the model fits the data if the frequency of the observed and the model-predicted individuals for each response pattern are close to one another (Finch & French, 2015, p. 189). To analyze the model fit, we used the bootstrap chi-square procedure in R program (whether the model fits for the overall data). The bootstrap chi-square test of overall model fit for a Rasch model was conducted by command *GoF.rasch(model.rasch, B=1000)*. First, the researchers analyzed the model fits for all items (50 items). The result shows that p-value is 0.006. If the p-value is less than 0.05, it means that the model does not fit the data. Thus, it is said that the model did not fit the data (for all items). Then the items fit model was analyzed (whether the model fits for the individual items as well) by command *item.fit(model.rasch, simulate.p.value = TRUE)*. There were three items that did not fit the model. These items are item number 27, 32, and 35. The data for these three items were removed, and the researchers analyzed the model which fits the data again.

The second analysis of the model fit of the data was done, and we got the p-value 0.017. It was still less than 0.05. It means that the Rasch model did not fit the data. Then the researchers analyzed the items fit the model for these 47 items. They got that the items number 3, 11, 13, 36, and 48 did not fit the model. The data for these items were then removed. Then, the researchers reanalyzed the model fit of the data with 43 items remained. The third analyzing of the model fit of the data showed that the model fits the data. It could be seen from the p-value were 0.053 (more than 0.05). Finally, after three times calibration of the fit-model, the researchers got the Rasch model fits the data without the items number3, 11, 13, 27, 32, 35, 36, and 48 (there are 42 items that would be analyzed). In other words, the researchers had gotten the overall model-fit for the Rasch model, then, they could continue the other analysis.

The researchers analyzed the difficulty level of the items, and the items fit the model. The summary of the analysis is clearly presented in Table 1.

The center of item difficulty level is 0; negative value represents relatively easy, and positive value indicates relatively more difficult items (Finch & French, 2015, p. 184). Based on that statement, it indicates that when the value of difficulty is increasingly negative, then the difficulty level of the problem is easier and when the value of the difficulty becomes more positive then the level of difficulty becomes increasingly difficult.

From the Rasch's analysis of the difficulty level of the items, it is found that the easiest question is item number 20 (with difficulty level -15.7892) and the hardest problem is item number 23 (with difficulty level 0.9702).

In theory, the difficulty levels are in the range of minus infinity to infinity. There are some items that have a good category based on their difficulty level. There are 28 good items, and the rest, 14 items, are not good based on the difficulty level. The not good items based on difficulty level are item number 5, 6, 7, 12, 14, 16, 17, 18, 19, 20, 25,29, 38,and 46. There are 69.77% of 43 items that are good in the difficulty level. Hence, the test in the moderate category based on the difficulty level.

Table 1. Difficulty level of items and the items fit of the model

Item No.	Difficulty level of the items	The items fit of the model
1	-0.8355	0.0792
2	-1.0570	0.6634
4	-0.3796	0.4554
5	-4.6802*	0.7165
6	-4.6802*	0.5149
7	-3.5262*	0.3861
8	-1.0317	0.1683
9	-2.8874	0.2574
10	-1.5902	0.3366
12	-5.0950*	0.6832
14	-5.7976*	0.6436
15	-2.6885	0.9208
16	-4.1508*	0.3465
17	-3.7959*	0.0891
18	-3.9593*	0.9208
19	-5.7976*	0.1584
20	-16.0705*	0.1881
21	-0.2267	0.0396#
22	-0.5127	0.9802
23	0.9695	0.3960
24	-1.8959	0.8713
25	-4.3832*	0.8614
26	-1.3516	0.7426
28	-1.7202	0.9604
29	-3.1221*	0.0693
30	-1.5902	0.2970
31	-0.4016	0.4356
33	-2.6282	0.4059
34	-1.7202	0.4653
37	-1.9713	0.9406
38	-3.5263*	0.3168
39	-2.0908	0.1287
40	-2.0102	0.1386
41	-1.1084	0.1386
42	-1.5589	0.2277
43	-1.4678	0.2277
44	-2.0908	0.8119
45	-2.9610	0.3762
46	-3.3073*	0.6436
47	-2.1756	0.4158
49	-1.3235	0.9505
50	-1.6541	0.3366

Notes:
*item is not good based on the difficulty level
#item misfit with the Rasch model

The teacher should pay attention to the not good category items. All of the items that are not good based on the difficulty level are categorized at too easy items. These items are not good because they are too easy for every examinee. It was indicated by all of their indexes of difficulty level which are smaller than -3.0.

Rasch model had fit with the data, but there is one item that did not fit with the Rasch model. This item is item number 21. We could not decide on these items. It was because these items did not fit with the model. It means that the characteristics of this item (item no. 21) based on the Rasch model were not adequately accurate.

The analysis of item characteristics is displayed in the form of curves for all items can be seen in Figure 1. The item characteristic curve (ICC) places the test participant's location on the latent trait measured on the x-axis and the ability to master an item on the y-axis (Finch & French, 2015, p. 184). The latent trait refers to the Javanese language ability, and the ability to master an item (probability answer correctly) refers to the probability of the examinee to respond correctly to the item. From ICC, it can be known about the probability of correctly answer from someone with a certain ability on an item. The command to get ICC for all items (42 items) together is *plot(model.rasch,type=c('ICC'))*. It gives us all the ICC of the item in the test.

Figure 1 shows the ICC of 43 items. It was difficult to interpret the curve if we used all ICC together. The ICC of the items number 23 was located at the most right position of the x-axis (Finch & French, 2015, p. 185). It means that the item number 23 is the most difficult item. The easiest item was not able to find, because it was so complex. However, it is clear that the item number 20 is the easiest item based on the difficulty level of the item. If the curve from these items is separated, we can see it more clearly. Thus, the ICC for item number 20, 23, and two other numbers can be compared. The ICC for item number 20 and 23, and two other items are presented in Figure 2.

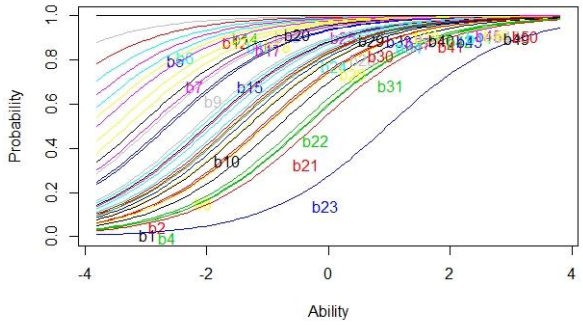


Figure 1. The ICC of Javanese language test

From Figure 1, some of ICCs are not good because the correct response probability for the examinee with low ability is high. These items are item number 5, 6, 7, 12, 14, 16, 17, 18, 19, 20, 25, 29, 38, and also 46 (total of 14 items). All of these items have fitted the model.

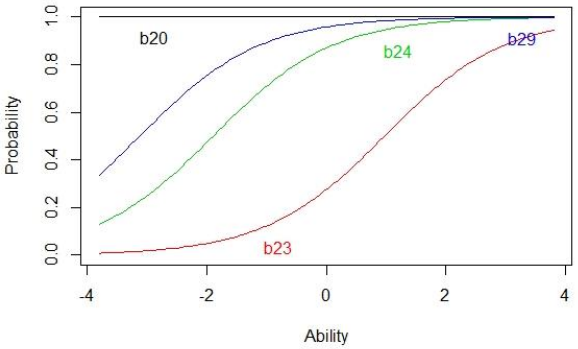


Figure 2. The ICC for items number 20, 23, 24, and 29

However, the difficulty levels of these items are not good. Thus, these items (see Figure 3) are not good based on the ICC and difficulty level.

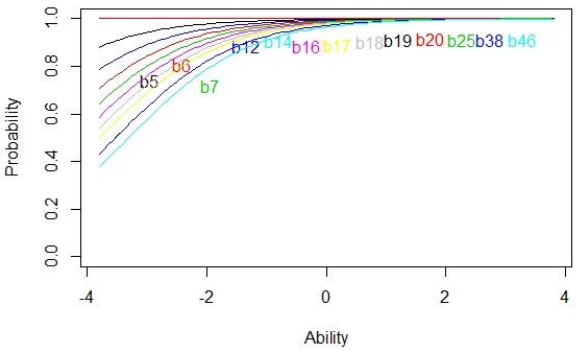


Figure 3. Items with not good ICC

To look the ICC of a specific item, let us say that items number 20, 23, 28, and 29,

used the command *plot(model.rasch,type=c("ICC "),items=c(17,20,21,25))*. It is a little different from the command for all ICC, in which, sort number from specific items was mentioned. It would make every ICC of some items in one graph to be able to compare easily.

Figure 2 presents some item characteristic information. For item number 20, regardless of the student's ability, the probability to answer correctly is the same for all examinee, which is 1.0 (always true). It indicates that the item number 20 is too easy for every examinee. It means that examinee with any Javanese language ability will be able to respond the item correctly (the examinee with ability value -4 through 4 could respond to this item correctly). For the hardest item (item number 23), the examinee with ability 1 will have probability approximately 0.5 to answer this item correctly. To get high probability about 0.9 or more, the examinee should have Javanese language ability almost 4. The Javanese language ability would be needed to increase the opportunity to answer this item correctly.

The test characteristic in correlating the ability with true score can be found by TCC (Test Characteristic Curve). True score is the sum of correct answer probability. The Javanese language test TCC is shown in Figure 4.

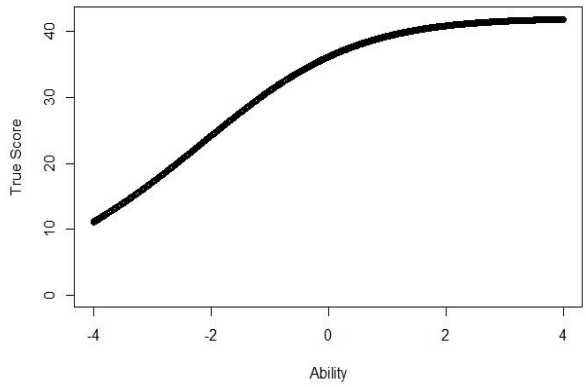


Figure 4. The TCC of the test

From Figure 4, it is known that the test is an easy category. The examinee with a low ability (-3) will have true scores approximately 19, and the examinee with an average ability (0) will have true scores approximately 35 (near to the maximum true score, that is 42).

The examinee with ability value 0 (average ability) will have a different probability for each item. He/she will have probability 0.2 for item number 23, probability approximately 0.8 or more for item number 24 and 29, and probability 1.0 (true response) for item number 20. Figure 2 explains that the difficulty level of item number 20 is easier than item number 24 and 29, and item number 24 and 29 are easier than item number 23. Figure 1 shows that some ICCs are not good since the correct response probability for examinee with low ability is high. These items are item number 5, 6, 7, 12, 14, 16, 17, 18, 19, 20, 25, 29, 38, and 46 (14 items). Those items have fitted the model. The item characteristic for every item can be described the same way as we had done to the item number 20, 23, 24, and 29, by separating it from the other ICC so that it will be seen clearly.

In addition to the ICC, we used the R program to plot the item information curve (IIC). The IIC describe the information function of an item. It refers to the degree to which item reduces the uncertainty in the estimation of Javanese language ability (the latent trait) value for an individual (Finch & French, 2015, p. 185). A high value of information for a specific range of ability distribution indicates that the item provides relatively more information regarding the latent trait (Javanese language ability) in that region than another region in the distribution (Finch & French, 2015, p. 186). Based on the IIC, we can see how reliable the item in giving information. All the IIC are shown in Figure 5. There are 50 IIC with each degree in estimating the information given by each item. The command to get IIC for all item in the test is *plot(model.rasch,type=c('IIC'))*. The command for specific IIC is *plot(model.rasch,type=c('IIC'),items=c(18,21,25,40))*, that will produce IIC for item number 20, 23, 28, and 47. The IIC for 43 items is shown in Figure 3, and the IIC for item number 20, 23, 28, and 47 is shown in Figure 7.

There are 43 IIC that can describe how reliable each item in the giving information about the Javanese language ability value for an individual. There are just 43 IIC of the 43 items that the Rasch model fits for the data. From Figure 4, we can get the most accurate and inaccurate items in giving information about the examinee's ability in the Javanese language. These are shown by item number

20 and 23. The IIC for these numbers is shown separately from the others in Figure 5 with item number 28 and 47.

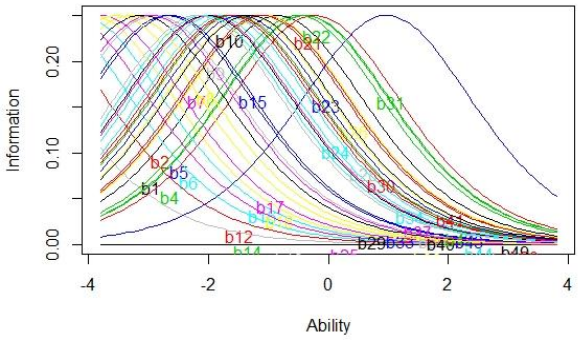


Figure 5. The IIC of 42 items

Some of IICs give maximum information for examinee with a low ability (Figure 6). These items are item number 5, 6, 7, 12, 14, 16, 17, 18, 19, 20, 25, 29, 38, and 46 (14 items). These items did not give maximum or give low information for the examinee with the medium or high ability. These items are not good, because they give maximum or high information just for low ability examinee and these items based on the ICC and the difficulty levels are not good. Therefore, we can conclude that these items are not good based on the ICC, IIC, and difficulty level.

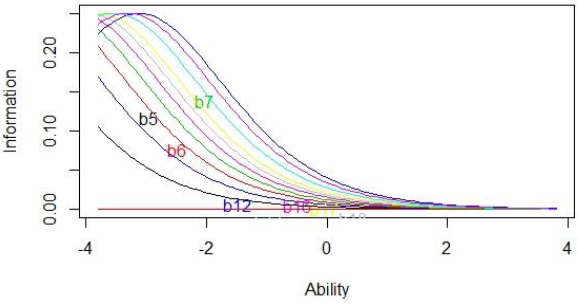


Figure 6. Item with not good IIC

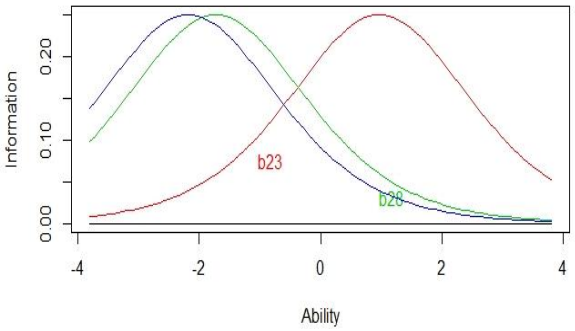


Figure 7. The IIC for item number 20, 23, 28, and 47

Figure 7 shows the IIC for item number 20 is the most inaccurate in giving information about the examinee's Javanese language ability. This item cannot give the information accurately because any examinee with any ability shows 0 information value that can be provided by this item. We cannot differentiate the examinee's ability. There is no information about the examinee ability (in the Javanese language) that we can get if we use this item to measure them. The IIC for item number 23 shows that it is needed ability approximately 1 to get information about 0.25, in other words that item 23 provides maximum information for estimating θ (Javanese language ability) around values of 1. The item number 28 and 47 will give maximum information about the examinee if he/she has ability about -2. The IIC for every item is different, but this study shows more specific item information curve for item number 20, 23, 28, and 47. If we want to look at the IIC from the other item, we can separate it from the others.

Item information curves show the information function for every item in the test. For the total information, the function can get from Test Information Function. There are some features of the test information function. These are defined for a set of the test items at each point on the ability scale, the amount of the information is influenced by quality and number of test items, etc. One of the most important features of the test information function is that the contribution of each item to complete information is additive (Hambleton & Swaminathan, 1985, p. 104). The test information curve that shows the total information function is like Figure 8. The command to get test information curve is *plot(model.rasch,type=c("IIC"), items=c(0))*.

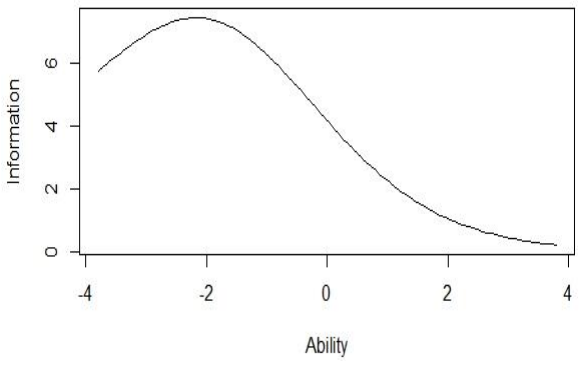


Figure 8. Test information curve

Figure 8 shows the estimate of the test information function on the curve. TIC presents how reliable the Javanese language test is. The TIC interpretation is similar to the IIC interpretation. The test provides us maximum information for estimating θ around values of -2. Thus, the test will be good to be used for examinee with low Javanese language ability. The test was less accurate in giving information on examinee with Javanese language ability 0 (average ability) or more than 0 ability.

The information function (IIC or TIC) has some application in the test construction, item selection, measurement precision assessment, test comparison, scoring weight determination, and scoring methods comparison (Hambleton & Swaminathan, 1985, p. 101). In item selection, we can select the item that can provide accurate information on examinee's ability. The item's IIC, which does not provide information, means the item should not be used in the test (like item number 20). The item does not provide information in any theta (ability), so it should not be used in the test.

Table 2. The information of each item in theta -3.0 until 3.0

Item No.	Information	Percentage
1	0.88	87.60%
2	0.86	85.78%
4	0.90	89.93%
5	0.16	15.74%
6	0.16	15.66%
7	0.37	36.94%
8	0.86	86.01%
9	0.53	52.58%
10	0.79	79.35%
12	0.11	11.01%
14	0.06	5.82%
15	0.57	57.43%
16	0.24	24.03%
17	0.31	31.04%
18	0.28	27.67%
19	0.06	5.82%
20	0	0.09%
21	0.90	90.31%
22	0.89	89.44%
23	0.87	86.55%
24	0.74	74.38%
25	0.2	20.06%
26	0.83	82.61%
28	0.77	77.38%
29	0.47	46.78%
30	0.79	79.39%
31	0.9	89.86%
33	0.59	58.87%
34	0.77	77.38%
37	0.73	73.00%
38	0.37	37.05%
39	0.71	70.70%
40	0.72	72.27%
41	0.85	85.29%
42	0.8	79.81%
43	0.81	81.09%
44	0.71	70.70%
45	0.51	50.76%
46	0.42	42.14%
47	0.69	68.98%
49	0.83	82.95%
50	0.78	78.42%

The complete information of the test across all values of the Javanese language ability (latent trait) can be obtained by using the command *information(model.rasch, c(-10,10))*. The subcommand c(-10, 10) identifies the range of the theta (ability) for which information is requested. The total information that is provided by the test at the examinee's ability ranges from -10 to 10 equal to 41.93 or 100%. It means that the test will give maximum information if the test were used in the examinees with ability -10 until 10. If we request for the ability values in range 0 to 10, with the command *information(model.rasch, c(0,10))*, is 5.9 or 14.08% of the total information provided by the Javanese language test. In the normal distribution raw, the area of range -3 to 3 equals to 95% of the total area. The total information that could be given by the test if we measure in the ability range of -3 to 3 is 24.98 or 59.58% of the total information. There is still moderate information which we could obtain by using this instrument in measuring the examinee with the ability in this range.

Beside the ICC, TIC, and the total information, we can get the information given by each item in the range of a certain ability (theta). In this study, the information, that is given by each item in the ability range of –3 until 3, are listed in Table 2. We can know the percentage that we get from the total information of each item.

Based on Table 2, we can see the information given by each item in the theta -3.0 until 3.0. The information can be used for item selection. How reliable the item depends on the percentage of information gotten from each item in this range of theta. We can set the criteria for reliable item like we need. For example, if we will compose a test, we cannot use item number 20, because it gives us very small information. If we set the criteria for reliable information of each item by more than 50%, we get 28 reliable items of 42 items that can be used (there are 66.67%). The remaining unreliable items (14 items) are not good. Incidentally, these unreliable items are also categorized as not good based on the ICC, IIC, and difficulty level.

Obtaining latent trait (Javanese language ability) estimates for the Rasch model in R program, we used the command *theta.rasch<-factor.scores.rasch(model.rasch)* to save the θ estimates from the Rasch model. Then, we used the command *summary(theta.rasch$score.dat$z1)* to get a basic descriptive statistic of ability(θ). The output of this command is shown in Table 3.

Table 3. The latent trait estimates

Min.	Median	Mean	Max
-2.0780	-0.1534	-0.1138	1.6538

We can see that the mean of Javanese language ability for the sample is -0.1138, with the minimum being -2.0780 and the maximal being 1.6538. The standard deviation of Javanese language ability gotten by the command *sqrt(var(theta.rasch$score.dat$z1))*. The result of the standard deviation of Javanese language ability is 0.750783. The plot of the latent trait (Javanese language ability) was gotten by the command *plot(theta.rasch)*. The plot of the latent trait (Javanese language ability) based on the Rasch model is shown in Figure 9.

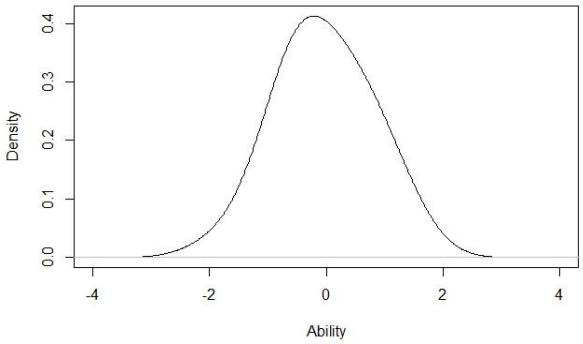


Figure 9. Plot of theta

Figure 9 shows that the distribution of Javanese language ability almost centered at 0. The center of the plot ability shows the mean of ability, that is -0.1367. Thus, that is the reason why it is almost centered for those with Javanese language ability value of 0. The highest density of Javanese language ability is located in the mean ability value. The distribution of the theta (Javanese language ability) based on the analysis using the Rasch model in R program shows the normal distribution

curve. The right side and the left side of the distribution curve are almost balanced.

Figure 8 shows that maximum information will be obtained when the Javanese language ability value is -2. However, the mean ability from the examinees is -0.1367, meaning that generally, the Javanese language test did not give maximum information on the examinee's Javanese language ability. It can be said that the test is less accurate. Thus, evaluation of the Javanese language test is needed.

The evaluation of the Javanese language test will make the test better, so that it can give more accurate information for a teacher in the assessment of precision measurement. The teacher will have further steps or ideas to be applied in the next Javanese language lesson if they know the examinee's ability generally to make the examinee's Javanese language ability increase. It is hoped that, with the increasing of the Javanese language ability, the student will practice it in their daily life. They retain the culture and character of Javanese language in their lives, which there are so much positive learning, culture, character, interaction in Java, and so much more.

This study analyzed the Javanese language test based on the Rasch model in the R program. For the next study, we hope they can use the other model to analyze the Javanese language test based on the procedure for each model. It is hoped there will be more test analysis, maybe about mathematics test, a certain language test, or the other test, especially the Javanese language test. Therefore, it will give the teacher a view to making a better test in the next chance that gives accurate information about the examinee ability and measures the examinee ability more accurate. It is better to use item response theory to analyze the test because there are some benefits that we can get. We can know about each item characteristic, the information function of each item, and the other benefits.

Conclusion

Based on the result of the analysis of Javanese language test using the Rasch model in R program, the interpretation, and the discussion, the researchers can conclude some points of the characteristic of the Javanese language test. The calibration of the fit-model was done in three times. It was done to get model fits the data with 42 items in the fit-model. Analysis of the difficulty level shows that there are 28 items of 42 items (66.67% of 43 items) that are a good category. Therefore, the Javanese language test is in the moderate category based on the difficulty level.

We can see the characteristic of the item in predicting the true probability for examinee with a certain ability in the ICC and the test characteristic from the TCC. Based on ICC and IIC, there are 28 good items (66.67%). Based on the information that we can get from each item (item information) in the theta -3.0 to 3.0, there are 28 items (66.67% give information more than 50%) of 42 items can be used (moderate category based on the information in this range of theta). From descriptive statistic, it can be said that the ability of examinees are in the moderate category because the mean of ability is -0.1138 (near from 0.00/average ability). Generally, the Javanese language test is in the moderate category. It will be better if we evaluate the Javanese language test to make a better test that gives more accurate information on the examinees' ability. The evaluation of the Javanese language test can be used by the Javanese language teachers to plan the next learning in their class to get better Javanese language learning.

Acknowledgment

The researchers thank Depok 1 Vocational High School, which had permitted the researchers to collect the data. Gratitude is also sent to contributors, Ali, Desy, and Laras for their help during the data collection and dealing with the dichotomy data tabulation.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Montery, CA: Cole Publishing.

Arikunto, S. (2010). *Prosedur penelitian: Suatu pendekatan praktik* (Revised ed). Jakarta: Rineka Cipta. Retrieved from https://doi.org/10.1017/CBO9781107415324.004

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London and New York, NY: Routledge Falmer.

Dede, C. (2010). Comparing frameworks for 21st century skills. In J. Bellance & R. Brandt (Eds.), *21st century skills: Rethinking how students learn* (pp. 51–76). Bloomington, IN: Solution Tree Press.

Downing, S. M. (2003). Item response theory: Applications of modern test theory in medical education. *Medical Education*, *37*(8), 739–745. https://doi.org/10.1046/j.1365-2923.2003.01587.x

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists: Multivariate applications book series*. London: Lawrence Erlbaum Associates.

Essen, C. B., Idaka, I. E., & Metibemu, M. A. (2017). Item level diagnostics and model - data fit in item response theory (IRT) using BILOG - MG v3.0 and IRTPRO v3.0 programmes. *Global Journal of Educational Research*, *16*(2), 87–94. https://doi.org/10.4314/gjedr.v16i2.2

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. New York, NY: Taylor & Francis.

Gregory, R. J. (2015). *Psychological testing: History, principles, and applications* (7th ed.). New York, NY: Pearson Education.

Gruijter, D. N. M., & van der Kamp, L. J. T. (2008). *Statistical test theory for the behavioral sciences*. New York, NY: Taylor & Francis Group.

Hailaya, W., Alagumalai, S., & Ben, F. (2014). Examining the utility of Assessment Literacy Inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education*, *58*(3), 297–317. https://doi.org/10.1177/0004944114542984

Hambleton, R K, & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.

Hambleton, Ronald K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Hambleton, Ronald K, & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Nijhoff.

Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, *7*(1), 1–19. https://doi.org/10.1038/s41598-017-03586-z

Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(1), 12–23. https://doi.org/10.21831/pep.v22i1.15609

Jambulingam, T., Schellhorn, C., & Sharma, R. (2016). Using a Rasch model to rank big pharmaceutical firms by financial performance. *Journal of Commercial Biotechnology*, *22*(1), 49–60. https://doi.org/10.5912/jcb734

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. (F. Mosteller, Ed.). Reading, MA: Addison-Wesley.

Mallinson. (2007). Rehabilitation institute of Chicago in rehabilitation research provides new insights. *Atlanta*, pp. 1–3.

Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan* (2nd ed.). Yogyakarta: Parama Publishing.

Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: SAGE Publications.

Purnama, D. N. (2017). Characteristics and equation of accounting vocational theory trial test items for vocational high schools by subject-matter teachers' forum. *REiD (Research and Evaluation in Education)*, *3*(2), 152–162. https://doi.org/10.21831/reid.v3i2.18121

Reckase, M. D. (2009). *Multidimensional item response theory (Statistics for social and behavioral sciences)*. New York, NY: Springer.

Setiawati, F. A., Izzaty, R. E., & Hidayat, V. (2018a). Analisis respons butir pada tes bakat skolastik. *Jurnal Psikologi*, *17*(1), 1–17. https://doi.org/10.14710/jp.17.1.1-17

Setiawati, F. A., Izzaty, R. E., & Hidayat, V. (2018b). Items parameters of the space-relations subtest using item response theory. *Data in Brief*, *19*, 1785–1793. https://doi.org/10.1016/j.dib.2018.06.061

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan*. Bandung: Trim Komunikata. Retrieved from https://umexpert.um.edu.my/file/publication/00013268_127390.pdf

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: Jossey-Bass.

van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*. New York, NY: Springer Science+Business Media. https://doi.org/10.1007/978-1-4757-2691-6 I

Young, D. J., Levy, F., Martin, N. C., & Hay, D. A. (2009). Attention deficit hyperactivity disorder: A Rasch analysis of the SWAN rating scale. *Child Psychiatry and Human Development*, *40*(4), 543–559. https://doi.org/10.1007/s10578-009-0143-z

Zubaidah, S. (2017). Keterampilan abad ke-21: Keterampilan yang diajarkan melalui pembelajaran. In *Isu-Isu Strategis Pembelajaran MIPA Abad 21*. Sintang, West Kalimantan: Program Studi Pendidikan Biologi STKIP Persada Khatulistiwa Sintang. Retrieved from https://www.researchgate.net/publication/318013627_keterampilan_abad_ke-21_keterampilan_yang_diajarkan_melalui_pembelajaran

REiD (Research and Evaluation in Education), 5(1), 2019, 75-84
Available online at: http://journal.uny.ac.id/index.php/reid

Developing an instrument to evaluate the influential factors of the success of local curriculum

Dedek Andrian
Faculty of Teacher Training and Education, Universitas Islam Riau
Jl. Kaharuddin Nasution 113, pekanbaru, Riau 28284, Indonesia
E-mail: dedek.andri321@gmail.com

Submitted: 12 March 2019 | Revised: 30 April 2019 | Accepted: 24 May 2019

Abstract

The purpose of this study is to create an instrument to evaluate the factors that influence the success of the local curriculum in senior high school. The instrument developed is essential to find out what factors influence the success of the local curriculum. This research is a research and development which consists of three stages: (1) initial investigation, (2) design and validation stages, (3) trials, evaluations, and revisions. The results of small scale trials were analyzed using EFA (Exploratory Factor Analysis), obtaining one invalid item from 32 items. The results of a wide-scale test analyzed using CFA (Confirmatory Factor Analysis) obtained all indicators have loading factors> 0.3 and reliability constructs> 0.7. The results of the analysis show that the developed instrument fits the data obtained in the field. From the constructs of the instrument development, it is indicated that there are nine influential factors, consisting of regional government support, local community support, local industry support, academic community support, student interest, student motivation, practical tools, practical materials, and funds to buy lab material. The instrument developed has been used properly to obtain information about the factors that influence the success of the local curriculum.

Keywords: *evaluation instrument, influential factors, local curriculum*

Permalink/DOI: https://doi.org/10.21831/reid.v5i1.23980

Introduction

Among all countries around the world, Indonesia is one well-known country with its various diversity in terms of tribes and cultures that exist in every region. This each-region uniqueness certainly needs to be appropriately maintained by every element of society and government. Each member of the community and government of the Republic of Indonesia needs to contribute to supporting sustainability for the country. The central and regional governments need to collaborate in providing stimulus to the community to create individuals who can support the sustainability of cultures in Indonesia. Thus, due to this large-scale diversity, maintaining harmony in it is an important thing that should be endlessly done.

One of the unique districts in Indonesia is the district of Meranti Islands. Meranti Islands is a National Food Security Development area which is well-known as the largest producer of sago in Indonesia, which has become the primary source for the community. There are 20% of Meranti Islands people who make sago as the main production for supporting their families. The Meranti Islands Regency community also has coconut, rubber, coffee, areca groves, which become the economic sources for the Meranti Islands Regency community as well (Central Bureau of Statistics, 2016). Besides, some people make financial resources from fishery products in the Regency. The regional potency which is owned by Meranti Islands Regency cannot yet be maximized as a means of creativity for the community in Meranti Islands Regency in im-

This is an open access article under the CC-BY-SA license.

proving the community's economy, because the population has not been able to manage the maximum yield of the plantations and fisheries. Most people only sell out the products from plantations and fisheries that have been harvested. Meranti Islands Regency people has not been able to utilize, preserve, and develop the independent natural resources they owned.

Based on the aforementioned problems, an effective solution is urgently needed so that the people in Meranti Islands Regency can make the region's potential as a source of creativity and can be the primary source of income for the people of Meranti Islands Regency in improving their financial condition. All elements of the community and government in Meranti Islands Regency need to think of strategic steps as an effort to preserve and develop the potential of the area in the Meranti Islands Regency. The central, regional, and all elements of society must strive to make adequate facilities and infrastructure that can function in developing natural resources or potential areas in Meranti Islands Regency.

The central government has designed efforts to develop and maximize the potential of the region in collaboration with local governments. Central and regional governments have made policies that are very effective in an organized system and called the local curriculum (LC). LC is a plan and arrangement regarding the content and material of the lesson as well as the method used as a guideline for learning activities set by the respective regions (Utomo, Sumiyati, & Suwandi, 1997, p. 1). This policy is regulated in the Regulation of Minister of National Education No. 22 of 2006 concerning content standards stating that local content and self-development activities are an integral part of the curriculum structure at the level of primary and secondary education. The basis of this policy is because the Republic of Indonesia consists of various regions with diverse geographical conditions, natural resources, and communities with different historical and cultural backgrounds.

The diversity above needs to be addressed with a system that can accommodate the needs of the regional government and the community in developing the potential of the region. It can be done through a local curriculum that is packaged in the form of learning activities. The school gets the authority from the government to develop the local curriculum by adjusting the potential of each region. Therefore, education units need to create and develop Graduates Competency Standards, Competency Standards, Basic Competencies, and other learning tools equipped with assessment tools and set minimum completeness criteria.

Decentralization provided by the central government to the government through the local curriculum gives full authority to local governments, schools, and communities to develop the potential of the region or natural resources in the Regions. Regions that have different potentials and natural resources can maximize the opportunities provided by the central government. Thus, the development of the local curriculum is handed over to the local governments to collaborate with schools and communities to develop curricula following the regional potential, socio-cultural, natural resources, and regional needs (Osborne, 2001, p. 660). Based on this policy, local governments that are more aware and understand their regional potential can develop and produce a product that can be sold so that it can improve the economy of the people in each region.

The LC is a bridge for regions to develop natural resources or regional potential. The LC becomes a system that is very effective in preserving and developing regional potential in each region. The LC designed through learning activities and run for two hours each week. Regional learning provides new colors for students in which students are directly involved in preserving and developing natural resources or regional potential (Prastiwi, 2013, p. 211). Preservation and development are applied directly by students in curriculum activities in the form of yoga, plantation, and entrepreneurship. The LC became a curricular activity favored by students in the Meranti Islands Regency, Riau Province. Students are very enthusiastic in participating in curriculum activities contained in the LC.

The LC becomes a facility and infrastructure that makes a very effective contribution in preserving and developing the potential of the region or natural resources owned by the Meranti Islands Regency. The LC is a breakthrough that guides local governments in bridging education with students, so that students can adapt to the environment, culture possessed by the Meranti Islands regency. Breakthroughs that have been designed by the central government through the LC make students able to develop and preserve the environment become something useful. Prastiwi (2013, p. 509) believes that it can help students recognize their cultural identity. In addition, through the content of the students, they can produce creativity for students, especially for students who do not continue their study at university.

The LC becomes a critical system in preserving the culture and potential of the region. Saputra (2013, p. 619) explains that LC is essential in maintaining the local wisdom of a region. LC can help the government in realizing the regional vision to enrich students with local cultural knowledge, student attitudes, and increase students' ability to socialize it to other regions (Prastiwi, 2013, p. 209). Therefore, the learning activities contained in the curriculum need to be carried out continuously in a classroom learning system. This learning activity is expected to provide maximum results in providing understanding and formation of student character so that the love for culture, natural resources, and potentials in the region will continually be growing.

Schools have sought preservation and development of natural resources or regional potential through the local curriculum (LC). Schools in Meranti Islands Regency have implemented local curriculum through weekly and theoretical learning activities. However, the LC that runs is not in line with the expectations. The LC which is expected to be an effective means to preserve and develop regional potential do not run optimally in accordance with procedures which are mandated by the ministry of education and culture. The implementation of regional learning in the local curriculum in schools does not have the objectives and targets to be achieved. Most teachers do not have the learning tools. The assessment carried out by the teacher is limited to the assessment of the cognitive aspect. As a result, the teacher has difficulty in assessing psychomotor aspects, for instance, the students' practice.

Kunter et al. (2013, p. 206) state that teacher education is an important variable as a quality of control that has a contribution in achieving the success of learning so that to achieve these goals, the teacher who teaches the subject must be in line with his field. Liakopoulou (2011, p. 66) insists that a teacher must have professional qualifications and knowledge as needed. However, high schools in the Meranti Islands Regency area do not have special local the LC teachers who administer the LC subject. The LC teachers who are in the school of the Meranti Islands Regency are from general subjects (Biology, Physics, Mathematics, economics, geography, etc.) because the school has difficulty in finding teachers who really have the LC expertise or teachers who have a background in cultural arts.

Evaluation is considered as an activity of identifying, clarifying, and also applying criteria of success to the results of a program (David, Kartowagiran, & Harjo, 2016, p. 28). Perez and Mardapi (2015, p. 149) state that the main purpose of the evaluation is to see the weaknesses or shortcomings of the education program. Evaluation can improve internal and external needs of educational programs such as curriculum, teachers, clarity of school programs in improving student learning achievement and available resources (Sugiyanta & Soenarto, 2016, p. 195). From the statement, it can be concluded that the role of evaluation is to help stakeholders make decisions based on the criteria they have set. Evaluation can be used to help stakeholders for many purposes in helping to improve the education process.

Curriculum evaluation is an integral activity of curriculum development activities. The curriculum that has been developed needs to be evaluated in order to see the achievement of the curriculum objectives that have been developed. However, curriculum

evaluation is not only used to obtain information about performance, but also as a measure of the value or effectiveness of any particular activity in education, whether it is a national project or every part of work done by the students. Hussain, Dogar, Azeem, and Shakoor (2011, p. 265) mention that 'Curriculum evaluation refers to information on the value and effectiveness of a particular program.' Bharvad (2010, p. 72) explains 'Curriculum evaluation refers to the process of studying the merit or worth of some aspect, of the whole a curriculum.' In other words, it can be interpreted that curriculum evaluation is in accordance with the process of awarding or value from several aspects or the whole of the curriculum.

Curriculum evaluation plays a very important role in getting information about the achievements of the curriculum that has been developed. Uys and Gwele (2005, p. 98) state that 'curriculum evaluation is a systematic way of examining all components of a curriculum whose results are in evaluative conclusions.' van den Akker and Verloop (1994, p. 422) state that 'curriculum evaluation is an important aspect that can answer fundamental questions about the achievement of the purpose of developing a curriculum.' From this information, it can be concluded through curriculum evaluation, evaluators, or stakeholders can check the achievement of the curriculum through basic questions concerning the objectives compiled.

Curriculum evaluation activities play an essential role in detecting the success of learning because the problems of learning outcomes that are not achieved are oftenly addressed in the curriculum. Yeung (2010, p. 190) argues that 'curriculum evaluation has an important role in solving problems that occur in the curriculum'. In addition, Hakan and Seval (2011, p. 593) agree that 'curriculum evaluation is a phase of choosing information, obtaining, analyzing, transferring, using, and making decisions to improve curriculum quality.' National Research Council (2004, p. 4) states that 'three components of the work frame must be determined to evaluate the curriculum, namely; (1) program materials and principles in designing curriculum, (2) quality of curriculum implementation, and (3) curriculum impact on student achievement. These three things can show how far the curriculum has been developed effectively.' Figueiredo, Leite, and Fernandes (2016, p. 283) suggest that curriculum evaluation is an activity of analyzing the process and impact of the curriculum that runs at school. Leathwood and Phillips (2000, p. 479) assert that with curriculum evaluation, other activities such as understanding, teaching practices, learning, and assessment can be improved. It shows that the curriculum is the basis for the success of teaching and learning activities conducted in the classroom.

According to Haghparast, Sedghizadeh, Shuler, Ferati, and Christersson (2007, p. 14), understanding of problems in learning can be known by investigating or evaluating a running curriculum. Schools can explore intensively approaches that can be used to improve the current curriculum (Adin-Surkis, 2015, p. 35). Thus, it can be concluded that the standard of the good or poor curriculum can be seen by evaluating the curriculum. Curriculum evaluation must be thoroughly evaluated because the curriculum developed is based on objectives, background, policy, and needs analysis equipped with infrastructure and supporting elements to achieve the objectives. All parts of the components that form the basis of the birth of a curriculum that starts from the component context, input components, process components, product components to the impact of the curriculum for individual students need to be evaluated.

Method

Research on LC input evaluation instruments was carried out using a research and development model. The purpose of this study was to produce a product in the form of an instrument to evaluate the input of the LC in a high school in the Meranti Islands Regency. The research model used in this study is the model compiled by Borg and Gall (1983) consisting of 10 stages which are simplified into three steps, namely: (1) initial investigation, (2) design and validation stages, (3) trials, evaluations, and revisions.

Initial Investigation Phase

The initial investigation in this study was carried out by a qualitative approach. Respondents or participants in this study were the curriculum principals and local culture teachers from four schools, namely: SMAN 3 Selatpanjang, SMAN 1 Tebing Tinggi Barat, MAN 1 Selatpanjang, and SMAN 2 Rangsang. The investigation was carried out by interviewing principals in the curriculum section and local culture teachers related to the supporting factors to succeed in implementing local curriculum at senior high schools in the Meranti Islands regency. Interviews in this research were used to explore what factors support the success of the LC in the senior high schools located in Meranti Islands Regency. The results of the interview were then analyzed using the method which was proposed by Miles and Huberman (1994), including; data reduction, data display, and also conclusion.

Model Design Phase

Based on the results of the initial investigation carried out through interviews with school principals in the curriculum and regional culture teachers, the main factors influencing the success of the local curriculum were obtained, namely; government support, community support, industry support, and support from the school's academic community, interests, student motivation, equipment, materials and funds to run the local curriculum. These factors are the key to the success of the local content curriculum that needs to be seen in the field using valid and reliable instruments.

Content Validity

The factors influencing the success of the local curriculum that have been obtained from the interviews were made in the form of questionnaires. Questionnaires that have been developed based on these factors were validated by experts. Expert validation was used to see how far the instrument content in the form of a questionnaire is valid and can be tested empirically in the field.

Trials

The trial was conducted twice, namely small-scale trials and large-scale trial. The small-scale trial aims to see the quality of the instrument in the form of a questionnaire that has been developed and the feasibility of the sample used. Small-scale trials were conducted on 100 high school students. The data obtained were analyzed using EFA. The large-scale trial also aims to see the quality of the construct of the instrument consisting of construct validity and construct reliability from the input factors used to evaluate the local curriculum. A large-scale test was conducted on 300 high school students. The results of the trial were analyzed by the CFA analysis technique with the help of the Lisrel 8.0 program.

Sources of Information

The sources of information in this study were the school principals in curriculum section, regional cultural teachers, and also students. The three sources of information are the subjects that clearly understand the local curriculum of senior high schools in the Meranti Islands Regency. This research was conducted at all high-school-level schools located in Meranti Islands Regency, Riau Province, Indonesia.

Findings and Discussion

Findings

Content Validity

The results of the expert measurement and evaluation were analyzed using the Aiken validity formula. The Aiken validity analysis is used to see whether the items of the instrument which was developed are in accordance with the indicators or factors which were found through the interview process. Aiken's validity would prove that the instruments that have been developed by the researchers are valid. Thus, the results of the analysis are clearly presented in Table 1.

Table 1. The Result of Aiken Analysis Validity

Factor	Item	Coefficient	Criteria
Government Support	1	0.778	Medium
	2	0.889	High
	3	0.889	High
	4	0.889	High
Society Support	5	0.778	Medium
	6	0.667	Medium
	7	0.889	High
	8	0.889	High
Neighborhood Industry Support	9	0.889	High
	10	0.778	Medium
	11	0.889	High
	12	0.778	Sedang
School Community Support	13	0.889	High
	14	0.889	High
	15	1.000	High
	16	0.889	High
Student Motivation	17	1.000	High
	18	0.889	High
	19	0.889	High
	20	0.778	Medium
Student Interest	21	1.000	High
	22	0.889	High
	23	0.667	Medium
Practice Tool	24	0.778	Medium
	25	0.778	Medium
	26	0.667	Medium
Practice Material	27	0.778	Medium
	28	0.667	Medium
	29	0.889	High
Practice Funds	30	0.778	Medium
	31	0.778	Medium
	32	0.778	Medium

Based on the analysis in Table 1, it can be concluded that of the 32 items developed, all of them are in the range of 0.4 to 0.8 and 0.8 to 1.00 with the categories 'Medium' and 'High'. These results indicate that the instruments developed can be used to evaluate the factors that influence the success of the LC.

Small-Scale Trial

Data from the small-scale trial were analyzed using EFA (Exploratory Factor Analysis) to see whether the number of samples used was feasible to be analyzed further by factor analysis. The analysis results show the value of Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) of 0.771 (KMO>0.5) and Bartlett's Test of Sphericity of 0.00 ($\alpha<0.05$). The results of this analysis indicate that the sample used as many as 100 respondents fulfilled the requirements to be analyzed using factor analysis. Furthermore, the anti-image correlation value was obtained 31 points of the anti-image correlation value which is greater than 0.5 and one item less than 0.5, namely item 31, so it can be concluded that the 32 items that have been developed, only 31 items that are worthy of being used to evaluate the factors that influence the success of the senior high school the LC in Meranti Islands Regency, while one other item must be discarded. The results of the analysis can be seen in Table 2.

Table 2. Anti image correlation

No	DP	No	DI	No	DM	No	DS	No	Mi	No	Mo	No	AP	No	BP	No	DP
	Anti Image		Anti Image		Anti Image		Anti Image		Anti Image		Anti Image		Anti Image		Anti Image		Anti Image
1	0.71	5	0.84	9	0.84	13	0.73	17	0.75	21	0.65	24	0.75	27	0.79	30	0.84
2	0.59	6	0.85	10	0.75	14	0.83	18	0.83	22	0.73	25	0.64	28	0.74	31	0.37
3	0.74	7	0.66	11	0.78	15	0.84	19	0.71	23	0.82	26	0.85	29	0.86	32	0.83
4	0.78	8	0.67	12	0.79	16	0.74	20	0.82								

Based on Table 2, of the 32 items which are developed, 31 items have a greater anti-image correlation value of 0.5, and are declared appropriate to be used to evaluate the input of the local curriculum so it can be concluded that only 31 items can be used to evaluate the LC. The following result is the reliability of the instrument using Cronbach's Alfa. The results of the analysis show that the input evaluation instrument that has been developed is reliable because it has the value of Cronbach's Alpha 0.82 (more than 0.70). The results of the analysis are clearly presented in Table 3.

Table 3. Reliability Statistics

Cronbach's Alpha	N of Items
0.820	32

Large–Scale Trial

A large-scale test with 200 respondents was used to see how far the indicators or factors influencing the success of local curriculum have an acceptable level of construct validity and reliability. The basic factors that lead to the success of the local content need to be tested to see how far these factors become the factors or indicators that influence the success of local content. This construct must be empirically tested and analyzed using the statistics of confirmatory factor analysis first order (CFA) to determine the level of validity and reliability of all constructs which are obtained from the interviews with deputy principals in the curriculum section. The results of the confirmatory factor analysis are presented in Table 4 and Table 5.

Analysis with the first order confirmatory factor analysis (CFA) also shows a model compatibility test of factors that influence the success of local content. The results of the CFA analysis indicate that the measurement model which has been formulated has Chi-Square value <2df, P-value> 0.05, and the RMSEA <0.08. GFI> 0.9 and AGFI> 0.9. The results of the first order CFA analysis indicate that the measurement model that has been formulated meets the provisions of the goodness of statistics.

Table 4. The result of confirmatory factor analysis

Factor	λ	Category
Government Support	0.60	Valid
Society Support	0.75	Valid
Neighborhood Industry Support	0.73	Valid
School Community Support	0.75	Valid
Student Motivation	0.66	Valid
Student Interest	0.65	Valid
Practice Tool	0.73	Valid
Practice Material	0.65	Valid
Practice Funds	0.45	Valid

Table 5. Construct reliability

Factor	λ	Error	Contruct Reliability
Government Support	0.60	0.64	
Society Support	0.75	0.44	
Neighborhood Industry Support	0.73	0.47	
School Community Support	0.75	0.44	
Student Motivation	0.66	0.56	0.87
Student Interest	0.65	0.57	
Practice Tool	0.73	0.47	
Practice Material	0.65	0.57	
Practice Funds	0.45	0.80	

The results of the construct reliability analysis in Table 5 show that the instruments that have been developed have acceptable reliability values (CR> 0.7). Based on the results

of the validity and reliability of the instruments developed, it can be concluded that the instruments that have been developed have met good validity and reliability so that it is appropriate to be used to evaluate the LC.

Discussion

The development of instruments to evaluate factors that influence the success of local content is carried out through a qualitative approach. Qualitative research can explore in depth what factors influence the success of a program through in-depth interviews with key participants who understand a program (Creswell, 2012). Quantitatively, these factors need to be detected using instruments that are valid and reliable so that the weaknesses or shortcomings of these factors can be corrected as early as possible. Valid and reliable instruments can provide accurate information about the weaknesses and strengths of an ongoing program or curriculum (Andrian, Kartowagiran, & Hadi, 2018; Hadi & Andrian, 2018). The instruments developed must be valid and reliable in content and construct because both components of validity and reliability are important points in the development of an instrument so that effective instruments can facilitate researchers to get good information (Tooth, Nielsen, & Armstrong, 2013). Effective instruments can fully describe what components or factors need to be improved to increase the quality of a program. Stakeholders can improve the program based on information from a measurement carried out by a valid, reliable, and effective instrument (Widodo & Sudarsono, 2016).

Conclusion and Suggestions

Conclusion

From the interviews, nine factors can influence the success of the local content curriculum, namely, local government support, local community support, local industry support, academic community support, student interest, student motivation, practical tools, practice materials, and funds to buy practice materials. The instrument has been validated by experts and analyzed using Aiken validity to produce valid items with the lowest coefficient value of 0.667 and the highest of 1.00. From the 32 instruments analyzed using exploratory factor analysis (EFA), only 31 instruments had an anti-image> 0.05, while one instrument that had an anti-image was less than 0.05. All indicators developed have a factor load value or more than 0.3 and the reliability coefficient is above 0.70 so that from the 32 items developed only 31 instruments can be used to evaluate factors that influence the success of the LC.

Suggestion

The instrument developed in this study portrays the factors that can influence the success of the local curriculum in the Meranti Islands Regency based on regional cultural characteristics so that further researchers are advised to develop the same instruments in areas with different characteristics. The development of the latest instruments is expected to provide in-depth insight into the development of instruments in certain contexts or characteristics.

References

Adin-Surkis, A. (2015). Teachers evaluate the new curriculum in English: Views regarding evaluation and evaluation tools. *Research in Education*, *93*(1), 34–59. https://doi.org/10.7227/RIE.0009

Andrian, D., Kartowagiran, B., & Hadi, S. (2018). The instrument development to evaluate local curriculum in Indonesia. *International Journal of Instruction*, *11*(4), 921–934. https://doi.org/10.12973/iji.2018.11458a

Bharvad, A. J. (2010). Curriculum evaluation. *International Research Journal*, *1*(12), 72–74.

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

Central Bureau of Statistics. (2016). *Produk domestik regional bruto Kabupaten Kepulauan Meranti menurut lapangan usaha 2011-2015*. Kepulauan Meranti: Badan Pusat Statistik Kabupaten Kepulauan Meranti.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

David, D., Kartowagiran, B., & Harjo, S. P. (2016). Evaluasi dan strategi pengembangan SMA Indonesisch Nerderlandsche School (INS) Kayutanam. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *20*(1), 27–44. https://doi.org/10.21831/pep.v20i1.7518

Figueiredo, C., Leite, C., & Fernandes, P. (2016). The curriculum in school external evaluation frameworks in Portugal and England. *Research in Comparative and International Education*, *11*(3), 282–297. https://doi.org/10.1177/1745499916661933

Hadi, S., & Andrian, D. (2018). Detecting teacher difficulties in implementing the local curriculum developed by the local government. *The New Educational Review*, *53*(3), 250–260. https://doi.org/10.15804/tner.2018.53.3.21

Haghparast, N., Sedghizadeh, P. P., Shuler, C. F., Ferati, D., & Christersson, C. (2007). Evaluation of student and faculty perceptions of the PBL curriculum at two dental schools from a student perspective: A cross-sectional survey. *European Journal of Dental Education*, *11*(1), 14–22. https://doi.org/10.1111/j.1600-0579.2007.00423.x

Hakan, K., & Seval, F. (2011). CIPP evaluation model scale: Development, reliability and validity. *Procedia - Social and Behavioral Sciences*, *15*, 592–599. https://doi.org/10.1016/J.SBSPRO.2011.03.146

Hussain, A., Dogar, A. H., Azeem, M., & Shakoor, A. (2011). *Evaluation of curriculum development process. International Journal of Humanities and Social Science* (Vol. 1). Retrieved from https://pdfs.semanticscholar.org/5fd2/070ac599fb07fc1e7146658e75add134c806.pdf

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583

Leathwood, C., & Phillips, D. (2000). Developing curriculum evaluation research in higher education: Process, politics and practicalities. *Higher Education*, *40*(3), 313–330. https://doi.org/10.1023/A:1004183527173

Liakopoulou, M. (2011). The professional competence of teachers: Which qualities, attitudes, skills and knowledge contribute to a teacher's effectiveness? *International Journal of Humanities and Social Science*, *1*(21), 66–78.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: SAGE Publication.

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11025

Osborne, E. W. (2001). Culture, development, and government: Reservations in India. *Economic Development and Cultural Change*, *49*(3), 659–685.

Perez, B. E. O., & Mardapi, D. (2015). Evaluation of the bridging course offered at a university to foreign students: Batches of 2012 and 2013. *REiD (Research and Evaluation in Education)*, *1*(2), 146–157. https://doi.org/10.21831/reid.v1i2.6667

Prastiwi, Y. (2013). Transmitting local cultural knowledge through English as Foreign Language (EFL) learning as a means of fostering "Unity in Diversity." *Academic Journal of Interdisciplinary Studies*, *2*(3), 507–513. https://doi.org/10.5901/ajis.2013.v2n3p507

Regulation of Minister of National Education, No. 22 of 2006 on Content Standard (2006). Republic of Indonesia.

Saputra, G. A. S. (2013). Enhancing local wisdom through local content of elementary school in Java, Indonesia. In *Proceeding of the Global Summit on Education* (pp. 614–620). Kuala Lumpur: World Conferences.

Sugiyanta, S., & Soenarto, S. (2016). An evaluation model of educational quality assurance at junior high schools. *REiD (Research and Evaluation in Education)*, *2*(2), 194–205. https://doi.org/10.21831/reid.v2i2.11118

Tooth, J.-A., Nielsen, S., & Armstrong, H. (2013). Coaching effectiveness survey instruments: Taking stock of measuring the immeasurable. *Coaching: An International Journal of Theory, Research and Practice*, *6*(2), 137–151. https://doi.org/10.1080/17521882.2013.802365

Utomo, E., Sumiyati, & Suwandi. (1997). *Pokok-pokok pengertian dan pelaksanaan kurikulum muatan lokal*. Jakarta: Departemen Pendidikan dan Kebudayaan.

Uys, L. R., & Gwele, N. S. (2005). *Curriculum development in nursing: Process and innovations*. New York, NY: Routledge.

van den Akker, J., & Verloop, N. (1994). Evaluation approaches and results in curriculum research and development in The Netherlands. *Studies in Educational Evaluation*, *20*(4), 421–436. https://doi.org/10.1016/0191-491X(94)00035-F

Widodo, E., & Sudarsono, F. X. (2016). Developing an observation instrument for assessing the effectiveness of English teaching at vocational secondary schools. *REiD (Research and Evaluation in Education)*, *2*(2), 135–154. https://doi.org/10.21831/reid.v2i2.8648

Yeung, S. S. (2010). Using school evaluation policy to effect curriculum change? A reflection on the SSE and ESR exercise in Hong Kong. *Educational Research Journal*, *25*(2), 117–209.

SUBMISSION GUIDELINES

- The manuscript submitted is a result of an empirical research or scientific assessment of an actual issue in the area of educational measurement, evaluation, and assessment 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 4,000-7,000 words (or 8-15 pages using the journal template) including the abstract, 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 implications, recommendations, or suggestions (if any); (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 are 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.
- 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 250 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 application.
- 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].
- All Author(s)' names and identity(es) must be completely embedded in the form filled in by the corresponding author: email; affiliation; and country. [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*