REiD (RESEARCH AND EVALUATION IN EDUCATION)
Vol. 3, No. 2, December 2017

Indexed in:
DOAJ DIRECTORY OF OPEN ACCESS JOURNALS · ISJD · Google Scholar · sinta

9 772460 699001

Research and Evaluation in Education

Vol. 3, No. 2, December 2017

Vol. 3, No. 2, December 2017 ISSN 2460-6995

Research and Evaluation in Education

Publisher:
PROGRAM PASCASARJANA
UNIVERSITAS NEGERI YOGYAKARTA

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

Publisher
Program Pascasarjana Universitas Negeri Yogyakarta

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

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

Foreword

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

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

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

Yogyakarta, December 2017

Editor in Chief

TABLE OF CONTENT

REiD (Research and Evaluation in Education), 3(2), 2017, 106-113
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

An evaluation of vocational high schools in Indonesia: A comparison between four-year and three-year programs

***[1]Soenarto; [2]†Muhammad Mustaghfirin Amin; [3]Kumaidi**
*Faculty of Engineering, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: soenarto@uny.ac.id

Submitted: 28 November 2017 | Revised: 15 December 2017 | Accepted: 22 December 2017

Abstract

The research aimed to gain insights into the quality of four-year program vocational high school (VHS) in Indonesia compared to four-year program VHS. This research was conducted based on the school graduate standard, business sector and industrial sector (or *Dunia Usaha dan Dunia Industri* (DUDI)) – or the performance of the graduates and alumni (the graduates' satisfaction). The research was conducted using Discrepancy Evaluation Model using 16 VHSs (eight four-year program VHSs and eight three-year program VHSs). The result shows that from the standpoint of the school, the graduates of the four-year program VHS are higher in quality than those of the three-year program VHS. The four-year program VHS graduates are more qualified in seven aspects: teamwork, discipline, tenacity, theoretical knowledge, confidence, creativity, and leadership. Meanwhile, using DUDI standpoint, the four-year program VHS graduates are also higher in quality than the three-year program VHS graduates. In addition, the four-year program VHS graduates are better in the quality of their discipline, tenacity, theoretical knowledge, practical skills, confidence, carefulness, creativity, and leadership. The four-year program VHS graduates have a higher level of satisfaction in terms of income than the three-year program VHS graduates. The higher quality of the four-year program VHS graduates has resulted from longer duration of the internship program (PKL) that provides them with reliable experience and skills concerning work-related problem-solving activities.

Keywords: *vocational high school, graduates, four-year program, three-year program*

How to cite item:

Soenarto, S., Amin, M., & Kumaidi, K. (2017). An evaluation of vocational high schools in Indonesia: A comparison between four-year and three-year programs. *REiD (Research and Evaluation in Education), 3*(2), 106-113. doi:http://dx.doi.org/10.21831/reid.v3i2.17077

Introduction

Education institution is a human resource production house with the managerial competency related to human resource and other related resources. Thus, it is the duty of education institutions to keep the process of improvement going and to produce graduates who fulfill the needs of the society. The society needs evolve as time changes and in alignment with the changes of circumstances. As *Asean Free Trade Area* (AFTA) and Asean Economic Society (MEA) were put into effect in 2003 and in 2015 respectively, the demand of business sector and industry sector (*dunia usaha dan dunia industri or DUDI*) for innovative and creative workforce is on the rise. On the other hand, free competition in the open market has made the distribution of goods, services, capital and market-ready skilled labor even more dynamic. To survive under such circumstances, Indonesia has to prepare itself

for upcoming chances and challenges in global market. Alisjahbana (2014) states that in the free trade era, from the standpoint of population, manpower and human resource, Indonesia has to pay more attention to three things: (1) keeping the demographic momentum, (2) improving the participation of manpower, and (3) improving the manpower productivity.

The afore-mentioned action of keeping demographic momentum is an action of keeping the advantage of Indonesian demography conducted by pushing the fertility rate and driving migration. Migration is an effective strategy to keep the economic growth. The demographic momentum as the foundation of Indonesian economic strength has to be followed by the effort to improve the manpower participation by nurturing flexible and efficient working climate and driving the participation of women in improving the national economy. The area of manpower participation is not the only area in need of improvement in Indonesia. Improvement is also needed in the area of manpower productivity, which can be a competitive advantadge that is able to improve the competitive edge of manpower in open market.

Vocational high school (VHS – or *SMK* (*Sekolah Menengah Kejuruan)* is one of the education institutions responsible for producing skilled workers with the ability to adapt to the changes in the need of the society as the effect of the dynamic international economy with the support of Indonesian demographic bonus. VHS can be a powerful weapon in improving manpower participation and productivity by taking advantages of education processes. Pardjono, Sugiyono, and Budiyono (2015) state that *'vocational education cannot be removed from existing workforce development'*. Under the same light, in their research, Ramayani, Aimon, and Anis (2012) conclude that Indonesian government has to support the efforts made to improve manpower productivity by producing policy that focuses on education and health and providing more fund in the area related to human resource building.

In the Law No. 20 of 2003 of Republic of Indonesia concerning National Education System, VHS is defined as the education institution responsible for preparing students to work in certain fields of work. Dewey (1916) argues that '*a vocation means nothing but such a direction of the life activities as renders them perceptibly significant to a person, because of the consequences they accomplish, and also usefull to his assocoate.*' Moreover, Thompson (1973) argues that vocational education improves students' skills that eventually will improve their productivity. VHSs then play an important role in determining the competitive edge of Indonesian manpower by providing ready-to-work and high quality workers for national and also international needs. As stated by Komariah (2010), vocational high school is an education institution responsible to prepare students for labor market and nation-building effort.

Prior to 1970, vocational high school and regular high school have the same study duration: three years. In 1970, as stipulated in through the First PELITA (Five-Year Building, or *Pembangunan Lima Tahun* in Indonesian term) Program, Indonesian government built eight four-year program engineering vocational high schools under the banner of '*Proyek Perintis Sekolah Teknologi Menengah Pembangunan*' ('Development Engineering High School Initiative Project'). In 1974, Indonesian government built four more four-year program vocational high schools – this time with agriculture as the concentration. The goal of this project is to prepare industrial technicians or workers with engineering skill possessing (1) initiative attitude, (2) ability to work and love the work, and (3) ability to understand, manage, and implement the ideas of engineers from upper level and to provide guidance to the technical workers from lower levels. The four-year program is expected to provide supports for vocational high schools in producing skilled workers with competitive edge. All of the goals of four-year program vocational high schools are in national education standard, specifically the standards for the graduates. In the Regulation of the Minister of Education and Culture No. 20 of 2016, the competence standard of the graduates (*Standar Kompetensi Lulusan*/SKL) is the formula of qualification criteria for graduates, which are achieved upon the completion of a series of programs and education in the area of attitude, knowledge, and skills.

In order to be able to achieve the goals of the Development Engineering High School Initiative Project – which is now known as four-year program vocational high schools – the Directorate of Vocational High School Administration focuses on the improvement of curiculum, learning and teaching process, and evaluation process. To take everything one step further, the Directorate also focuses on the improvement of teacher professionalism and builds cooperation with parties involved in business sector and industrial sector (DUDI). However, there were doubts related to the effectiveness of four-year program vocational high school as The National Statistics Board (*Badan Pusat Statistik/BPS*) released data related to the number of unemployment in Indonesia in 2014. The data show that there were 2.179 million unemployed graduates of vocational high schools, which is 15.15% of the total number of unemployment in Indonesia for above-15-year-old workforce. The number is an accumulation of all unemployed graduates of four-year program vocational high schools and the graduates of three-year program vocational high schools. There were no distinction made between the graduates of four-year program vocational high schools and those of three-year program vocational high schools in the data presented by BPS although both of them do not follow the same education process. This phenomenon then made us wonder about the quality of the graduates of both programs and the differences. Table 1 shows the number of open unemployment with VHS education.

Table 1. Vocational high school graduate unemployment data in 2011-2014

Year	Total Number
2011	2,270,873
2012	2,085,474
2013	2,122,850
2104	2,179,886

The questions related to the worth or merit of four goals of the four-year program VHSs can only be answered through evaluation. Stufflebeam and Shinkfield (1984) define evaluation as *'systematic assessment of the worth or merit of some objects'*. In this case, evaluation is conducted to define the worth or merit of the goals of four-year program vocational high schools. Stufflebeam, Madaus, and Kellaghan (2000) state that the process of evaluation should not be alien to the process of comparing. The evaluation of the worth or merit of the four-year program VHSs is conducted by comparing the competence of the graduates of both programs. The competence of the graduates is measured with the standards set by schools of origin as the provider of education services, the standards of *DUDI* (in terms of the performance of the graduates) as the employer of the graduates, and the personal standard of the graduates (level of satisfaction) related to their jobs.

Method

The goal of this evaluative research was to gain insights into the quality of education provided in both programs (three-year program and four-year program) of vocational high schools. The method applied in this research was Discrepancy Evaluation Model (DEM) developed by Provus. The Discrepancy Evaluation Model identifies discrepancy between the standards used as the basis of assessment and the performance in reality (Kaufman & Susan, 1982, p. 127). This research used three-year program vocational high schools' graduates as the basis of assessment. The performance of the graduates of the three-year program vocational high school was set to be the basis or standards of assessment because it was the basis of the innovation that was known later as four-year program vocational high school. Inovation in this case is the production of something better than the existing product or program.

This research was conducted in eight three-year program VHSs and eight four-year program VHSs. All of the selected four-year program VHSs were part of the early four-year program initiative. On the other hand, all of the selected three-year program VHSs were selected based on the similarities with the selected four-year program VHSs in terms of the area of the school location. The respondents included all parties involved in the management of the vocational high schools, such as (1) head master, (2) vice head master, (3)

head of skill programs, (4) labor market coordinator, (5) guidance and counseling coordinator, (6) alumni, and (7) business sector and industrial sector. Table 2 shows detailed information about the respondents involved in the research.

Table 2. Research respondents

No	Research Subjects	four-year program	three-year program	Total
1	Head Master	8	8	16
2	Vice Head Master	32	32	64
3	Head of Skill Program	40	40	80
4	Special Labor Market Coordinator	8	8	16
5	Guidance and Counseling Coordinator	8	8	16
6	Alumni	40	40	80
7	DUDI	40	40	80
	Total	**176**	**176**	**176**

The research data were collected using questionnaire, observation, interview and documentation. The questionnaire in the data collection process was distributed to reach the opinions from schools and parties in *DUDI* about the competence of the graduates from both programs. In addition, the distribution of the questionnaire was also conducted to gain insights into the level of satisfaction of the graduates related to their jobs. The result of the validity test using a questionnaire showed that the instrument used was capable of measuring the data validly. The instrument reliability estimation shows that the questionnaire had the reliability coefficient as much as 0.83 which can be categorized as reliable. In collecting the supporting data related to the graduates, this research used not only a questionnaire but also observation, interview and documentation. The validity of the interview guide and observation guide was tested by experts via expert judgement.

In order to make the data fit to be presented in the form of tables and diagrams, the collected data from various instruments were processed through tabulation and analysis process using the descriptive statistics. In this step, the qualitative data were projected as a support for quantitative descriptive findings.

Findings and Discussion

The result of the data analysis on the competence, performance, and level of satisfaction of the graduates of the four-year program vocational high schools in comparison with that of the graduates of the three-year program vocational high schools is described as follows.

The Graduates' Competence

There are 11 aspects studied in this research, including team-work, discipline, ethics, tenacity, theoretical knowledge, practical skill, confidence, carefulness, creativity, sense of responsibility, and leadership. Figure 1 shows the competence of the graduates of four-year program and three-year program vocational high schoosl from the standpoint of school.

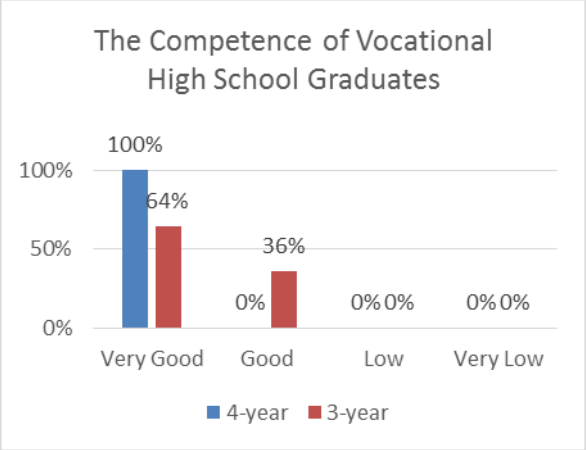


Figure 1. The competence of vocational high school graduates from the standpoint of school

Upon the analysis on the above-mentioned aspects, from the standpoint of school – represented by head masters and vice head masters – all of the graduates of four-year program vocational high schools possess 'very good' competence. This result is better than the result for graduates of three-year program vocational high schools in which only 64% of them are in 'very good' category and the rest (36%) are in 'good' category.

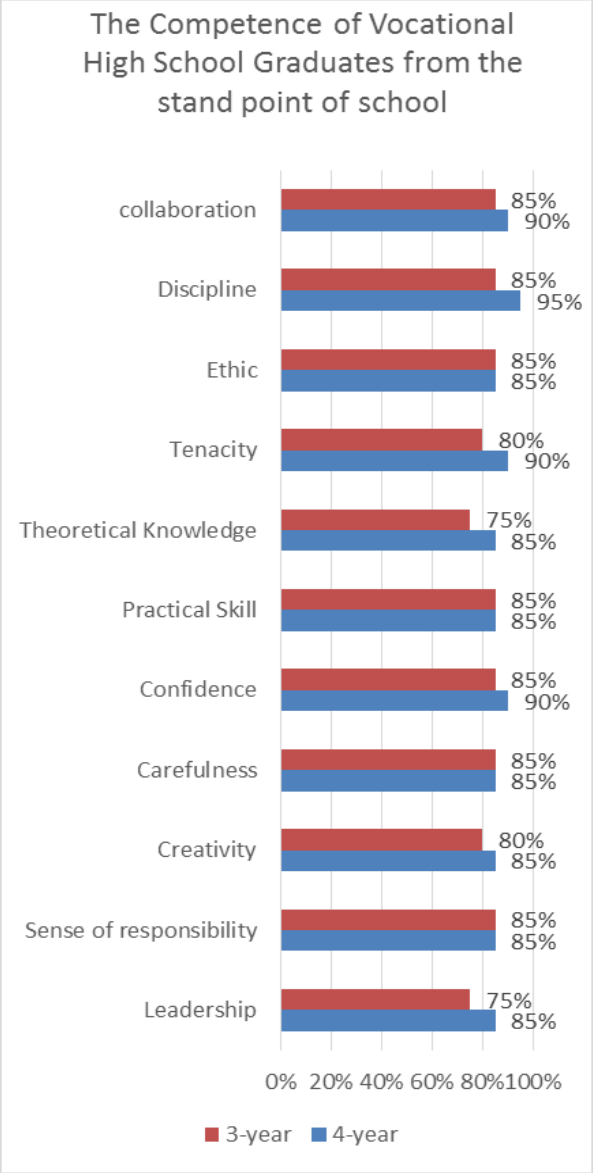


Figure 2. Aspects of competence of vocational high school graduates

The result of the assessment on all of the aspects shows that the graduates of four-year program vocational high schools are better in seven out of eleven assessed aspects (teamwork, discipline, tenacity, theoretical knowledge, confidence, creativity, and leadership). On the other four aspects, the competence of the graduates from both programs are in the same level. The superiority of the graduates of the four-year program vocational high schools on the seven aspects resulted from their rich experienced gained in a longer internship program (or *Praktik Kerja Lapangan (PKL)* in Indonesian term). This longer internship program facilitated the students of four-year program vocational high schools with proficient time for in-class knowledge internalization. In addition, the longer internship program made the students more experienced in problem-solving activities in the real daily work. The superiority of the four-year program vocational high school graduates in the seven aspects made them more competent at the business world. Figure 2 shows the competence of the graduates of both programs in every measured aspect.

The Performance of the Graduates

In this research, the performance of the graduates from both programs is also analyzed from the standpoint of business sector and industrial sector (*DUDI*), specifically their aspects of competence. There are 11 aspects measured, including teamwork, discipline, ethic, tenacity, theoretical knowledge, practical skills, confidence, carefulness, creativity, sense of responsibility, and leadership. Figure 3 shows the comparison in terms of performance of the graduates from both programs.

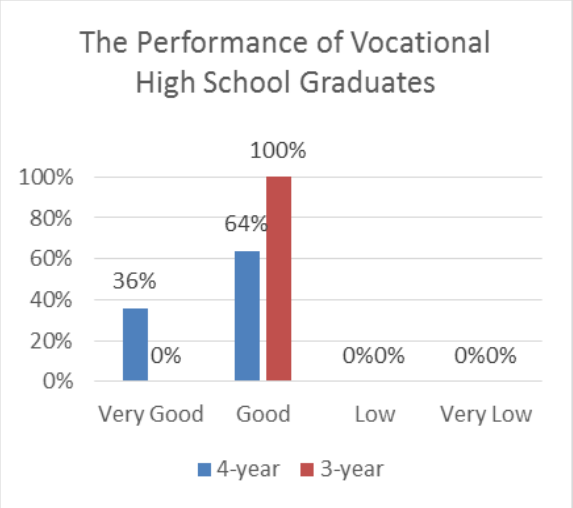


Figure 3. The performance of vocational high school graduates

The research result shows that 36% of the four-year program vocational high school graduates are in 'very good' category, whereas 64% of them are in 'good' category. The overall performance of the graduates of four-year program VHSs is better than that of the graduates of three-year program VHSs since all of the graduates of three-year program VHSs are in 'good' category .

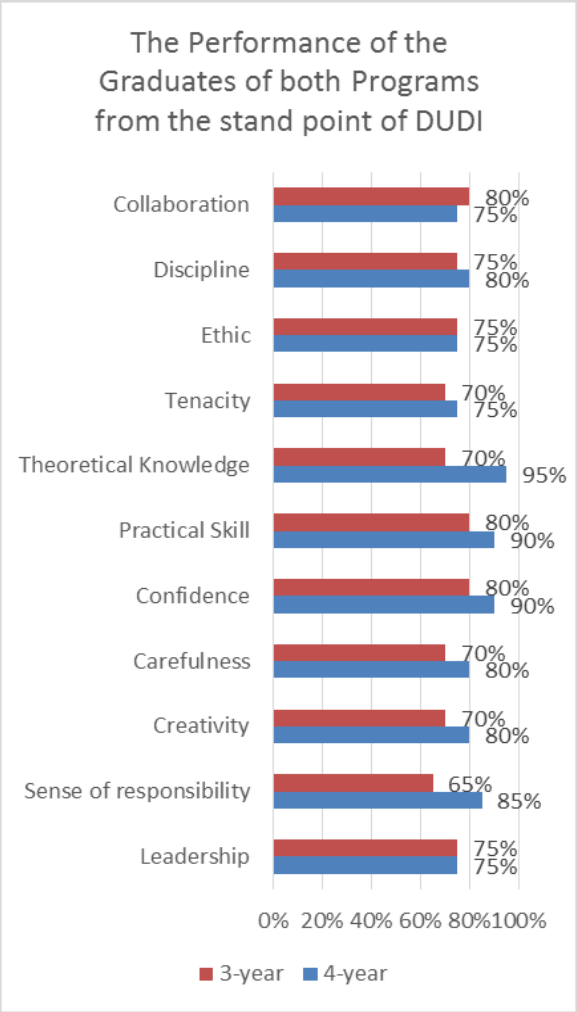


Figure 4. The performance of the graduates from the standpoint of DUDI

According to the statement of the employers or DUDI, the four-year program vocational high school graduates show superiority in eight aspects in terms of performance. They are superior in eight aspects - 72.72% of total aspects studied – including discipline, tenacity, theoretical skill, practical skill, confidence, carefulness, creativity, and also leadership (see Figure 4). As in the competence analysis from the standpoint of the schools, the superiority in these aspects is resulted from the longer internship programs which provide the students with richer and reliable experience. However, there is something new and intriguing in this competence analysis from the standpoint of DUDI. In the aspect of teamwork, the graduates of the four-year program vocational high schools are inferior to that of the three-year program vocational high schools. This is driven by the fact that the graduates of the four-year program vocational high school have the ability to accomplish tasks individually since they are equipped with higher level of competence and experience.

The Satisfaction of the Graduates

The satisfaction of the graduates is an accumulation the graduates' personal opinion about their jobs. There are nine aspects studied from this standpoint: income, working atmosphere, relationship with supervisors, relationship with co-workers, intention to get another job, working satisfaction, working facilities, working environment, and health insurance. The data of the satisfaction level of the graduates related to their jobs were collected by distributing questionnaires to the graduates of vocational high schools.

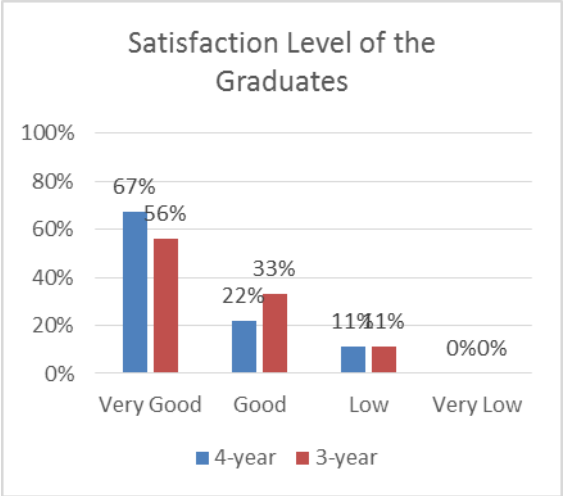


Figure 5. Satisfaction level of the graduates

Figure 5 shows the satisfaction level of the graduates of the both programs. The data show that 67% of the graduates of four-year program vocational high schools express 'very good' level of satisfaction toward their jobs. There are 22% of them who express 'good' satisfaction and 11% of them are in 'low' category. The level of satisfaction of the four-year program vocational high school graduates is higher than the level of satisfaction of the three-year program vocational high school graduates. There are 56% of the three-year program vocational high school graduates who express 'very good' level of satisfaction toward their jobs. As many as 33% of them are in 'good' category and 11% of them are in

'low' category. The most staggering difference is found in the aspect of income; the data show that the monthly income of the graduates of four-year program vocational high schools is between Rp 1,100,000.00 and Rp 5,000,000.00, while the income of the graduates of three-year program vocational high schools is between Rp 1,000,000.00 and Rp 2,500,000.00. The income of the graduates of four-year program vocational high schools is aligned with the level of competence and performance.

Conclusion and Suggestions

Conclusion

In conclusion, the result of the research can be concluded as follows: (1) from the standpoint of school, the competence of the graduates of four-year program vocational high schools is superior in seven aspects, including: teamwork, discipline, tenacity, theoretical knowledge, confidence, creativity, and leadership; (2) from the standpoint of the employers of the graduates (DUDI), the graduates of four-year program vocational high schools are superior in eight aspects: discipline, tenacity, theoretical knowledge, practical skill, confidence, carefulness, creativity, and leadership; (3) from the standpoint of personal satisfaction of the jobs obtained, the graduates of four-year program vocational high schools expressed a higher level of satisfaction, specifically in terms of incomes; (4) the competence superiority of the graduates of four-year program vocational high schools resulted from a longer internship program (PKL) which provided students with richer skills as well as experience related to problem-solving activities in real daily work.

Suggestions

Based on the conclusion, some suggestions are proposed as follows: (1) reconsider the role of Internship program in the students learning process. For optimized results, there should be more systematic, effective and efficient development, execution, and evaluation in the internship program of the four-year program vocational high schools; (2) for the internship program to be more efficient and effective, there should be unified evaluative efforts among all involved parties in schools and in business sector and industrial sector; (3) the result of the research shows that the graduates of the four-year program vocational high school are more superior than the graduates of the three-year program vocational high school in terms of working competence. Hence, there should be better appreciation for the graduates of the four-year program vocational high schools; (4) even though the graduates of the four-year program vocational high schools have sufficient skills to complete tasks presented individually, there should be more team-work-focused learning process for them.

References

Alisjahbana, A. S. (2014). Arah kebijakan dan program di bidang kependudukan, ketenagakerjaan, dan sumber daya manusia menghadapi globalisasi khususnya masyarakat ekonomi ASEAN. In *Seminar Nasional Tantangan Kependudukan, Ketenagakerjaan, dan SDM Indonesia Menghadapi Globalisasi Khususnya Masyarakat Ekonomi ASEAN*. Jakarta: Ikatan Praktisi dan Ahli Demografi Indonesia (IPADI).

Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Dover Publication.

Kaufman, R., & Susan, T. (1982). *Evaluation without fear*. London: New View Points.

Komariah, K. (2010). Memimpikan SMK di masa depan. In *Seminar Nasional Prospek Pengembangan Pendidikan Vokasional dalam Era Globalisasi* (pp. 127–132). Bandung: Culinary Education Study Program, FPTK, Universitas Pendidikan Indonesia.

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

Pardjono, P., Sugiyono, S., & Budiyono, A. (2015). Developing a model of competency certification test for vocational high school students. *REiD (Research and Evaluation in Education)*, *1*(2), 129–145.

https://doi.org/http://dx.doi.org/10.21831/reid.v1i2.6517

Ramayani, C., Aimon, H., & Anis, A. (2012). Analisis produktivitas tenaga kerja dan pertumbuhan ekonomi Indonesia. *Jurnal Kajian Ekonomi*, *1*(1). Retrieved from http://ejournal.unp.ac.id/index.php/ekonomi/article/view/738

Regulation of the Minister of Education and Culture No. 20 of 2016 on the competence standard of primary and secondary education graduates (2016). Republic of Indonesia.

Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (2000). *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.). Boston, MA: Kluwer Academic Publishers.

Stufflebeam, D. L., & Shinkfield, A. J. (1984). *Systematic evaluation: A self-instructional guide to theory and practice*. Dordrecht: Springer Netherlands.

Thompson, J. F. (1973). *Foundations of vocational education: Social and philosophical concepts*. Englewood Cliffs, NJ: Prentice-Hall.

REiD (Research and Evaluation in Education), 3(2), 2017, 114-123
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Developing physics problem-solving skill test for grade X students of senior high school

[1]Amipa Tri Yanti Nadapdap; *[2]Edi Istiyono
*Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: edi_istiyono@uny.ac.id

Submitted: 24 July 2017 | Revised: 29December 2017 | Accepted: 29 December 2017

Abstract

This research aimed to develop a physics problem-solving skill (PSS) test for grade X students of senior high school which met test instrument characteristics and feasibility. The development stages included: (a) test designing, (b) test trial, and (c) test revision and preparation. The designing stage included: (1) needs analysis, (2) mapping, (3) drawing conclusion, (4) determining test purpose, (5) determining competencies, (6) determining materials, (7) preparing answers, (8) writing items, (9) validating content, (10) improving and preparing the test, and (11) preparing the scoring guide with PCM. The trial stage consisted of: (1) determining trial subjects, (2) performing trial, and (3) analyzing trial result data based on IRT. The study was performed in Kulonprogo involving 281 students. The result shows that the instrument fulfills content validity with Aiken's V of 0.95 to 0.98. Based on INFIT MNSQ criteria, 52 items fit PCM, item difficulty index ranges from -1.47 to 0.88, meaning that all items are good, and information function analysis and SEM show that the test fits the ability between -1.3 and 2.7. Therefore, the test instrument meets the characteristics and feasibility to measure physics PSS in high school.

Keywords: *problem-solving skill, testing, physics, assessment*

How to cite item:

Nadapdap, A., & Istiyono, E. (2017). Developing physics problem-solving skill test for grade X students of senior high school. *REiD (Research and Evaluation in Education), 3*(2), 114-123. doi:http://dx.doi.org/10.21831/reid.v3i2.14982

Introduction

Assessment in education must be performed in order to measure student's cognitive skills. It is expected to increase the success of learning process. Thus, a series of test assessment instruments should be developed.

A test is a planned measurement instrument used by educators to give an opportunity to students to show their achievement and it is related to predetermined objectives (Cangelosi, 1995). A test can show the success rate of teaching based on target aspects. Its preparation is adjusted to its purpose, e.g. a summative test is used to measure student's achievement, formative test is to measure the success of learning process and a diagnostic test is to examine student's difficulty before a teaching and learning process.

There are other test types used to measure certain skills, such as cognitive, affective and psychomotor skills. A test has many variations in its preparation, i.e. multiple choice, sentence completion, listing, true-false, essay, matching, and modified form (Tonidandel, Quiñones, & Adams, 2002) Therefore, a test should be developed consistently, adjusted to its form and measurement purpose.

Problem solving is a skill which should be improved in the 21st century. Indonesia is a

developing country in terms of education, so problem-solving skill (PSS) is a skill which must be mastered by students in Curriculum 2013 (K-13). Rating in K-13 is done in the form of authentic assessment that assesses the start of the input, process and results (outputs) of learning, including attitudes, knowledge and skills. An assessment technique is relevant with the scientific learning process and able to assess the students' ability in the teaching and learning process and results. Regulation of Minister of Education and Culture No. 59 of 2014 states that problem-solving skill is required to achieve the objectives of K-13 to give students the life skills to be an individual and citizen who is faithful, productive, creative, innovative, and affective, as well as able to contribute to social life, nation, country, and world civilization. This skill is expected to produce scientific students (Nadapdap & Lede, 2016). Therefore, problem-solving skill test should be developed.

Problem-solving consists of four parts: (1) understanding a problem; (2) preparing a plan for solution; (3) performing a plan; (4) reexamining (Pólya, 1957). The indicators of problem-solving development according to Helaiya (2010) are including: (1) the ability to identify problem and problem-solving process; (2) the ability to define problem by thinking about different situations from the reality; (3) the ability to think of many possible alternatives of some solutions; (4) the ability to verify result of solution; and (5) the ability to verify in a solution acquisition process. Therefore, the aspects of a problem-solving test can be developed, including: (1) understanding; (2) planning a solution in problem solving; (3) describing a problem; (4) finding a way to solve a problem according to the planned solution; (5) bringing about a problem; and (6) evaluating the problem solving result assessment (Helaiya, 2010).

In physics teaching, PSS is the main topic in physics education research (PER) because it has long-term benefits. Further, physics PSS can help students understand the concepts of physics in real terms.

The most important part in teaching physics is students are expected to understand the real world. The theory of learning is based on one's process with its various interactions to gain experience which makes one have changes in cognitive, affective and psychomotor skills (Slameto, 2010).

According to Bloom, cognitive process thinking consists of Lower Order Thinking which consists of abilities to memorize, understand and apply, and Higher Order Thinking which consists of the ability to analyze, evaluate and create. PSS is a part of higher order thinking (Carvalho et al., 2015). Higher order thinking skills (HOTS) are: (1) higher order thinking at the upper part of Bloom's cognitive taxonomy, (2) teaching purpose behind cognitive taxonomy which can prepare students to perform knowledge transfer, (3) ablility to think, which means that students can apply the knowledge and skills that they develop during the learning process in a new context (Brookhart, 2010).

PSS can be measured by using a test which is consistent with the purpose of student's higher order thinking. Besides, the test which is used has to require the use of knowledge and skills in the new situation. In order to assess the HOTS, something new should be used. One of the ways to do that is using a test which is in the valid category — a test which is aimed to measure the HOTS.

One of the modern measurement theories is called Generalized Partial Credit Model (GPCM). GPCM is the improved Partial Credit Model (PCM). The PCM discriminant items are constant or 1, while the value GPCM discriminant varies. PCM is also appropriate for analyzing the response to the measurement of critical thinking and conceptual understanding in science (Istiyono, 2016). PCM was developed to analyze the test items that require several steps to resolve.

GPCM can be applied to tests, which is done with the steps that are clear for the testee. A physics achievement test is a test administered following the exact steps. Therefore, GPCM is expected to be applied properly.

Multiple-choice test has advantages, including: (1) the material being tested can cover most of the learning materials, (2) the students' answers can be corrected easily and quickly, and (3) the answers to each question is obviously right or wrong, so it is an objec-

tive assessment (Istiyono, 2016). Therefore, using a multiple-choice item test to measure the problem-solving skills is good to do.

Assessment in education uses two kinds of measurement theories: classical measurement theory and modern measurement theory or item response theory (IRT). The classical test theory (CTT) is also called the True-Score Classical Theory. The CTT is so named for the elements of this theory have been developed and applied for a long time, but still survive today (Suryabrata, 2002). According to the classical theory of measurement, measuring by using measurement score result is usually conducted partially based on the steps that must be taken in order to correct an answer items. Scoring is conducted at every step and score each item participant adds a score obtained by the students of each step, and the ability is estimated by the raw scores.

A scoring model is not necessarily right, because the level of difficulty of each step is not taken into account. Since a test is an instrument that provides stimulus in the form of a command or a question which requires a response from the test participants, the response which is given by the test participants stated in a score is easy to interpret.

In addition, the scoring results of a multiple choice test is gained by the use of a dichotomous model, which means that if the item response is correct, it is given a score of 1 and if the response is wrong, it is given a score of 0. Teachers do not use polytomous scoring models that would be more equitable because it considers item response measures. These dichotomous scoring models have yet to appreciate the steps of problem solving, because different error rates will result in the same score of 0. Dichotomous scoring models are certainly less fair. One of the scoring guidelines that can be selected is the provision of each category, as presented in Table 1.

HOTS is interdependent with students' problem-solving skill. Physics PSS can really help students solve physics problems in learning. With that skill, students are expected to solve a given problem with an effective solution. An accurate solution is seen based on the aspects to be measured, the aspects which measure students' problem-solving skill consistent with a students' operational stage of formal thinking. High school students are 17 years old in average, an age when they can think abstractly and logically which is categorized as problem solving stage.

Table 1. Scoring category & description

Category	Guidelines
Category -1	The students are wrong in writing the concepts used and the results are wrong. This is indicated by the students that answer question one and also one of the reasons
Category -2	The students are wrong in writing the the concepts used but the results can be correct. This is indicated by the students' correct answer to questions wrong basis.
Category -3	The students are correct in writing the concepts used but the end result is wrong. This is indicated by students' wrong answer to the question and correct reason.
Category -4	The students are correct in writing the concepts used and the results are correct. This is indicated by the students' correct answer to questions and correct reason.

(Istiyono, 2016)

Thinking skill is required in scientific thinking. Further, scientific thinking is involved in hipothetico-dedutive and inductive types (Piaget, 2005). Scientific thinking is working effectively and systematically, as well as proportionally. In terms of PSS, at that age, students can draw conclusions and interpret and develop hypotheses.

However, the existing test did not describe the skill which demands thinking consistent with the optimization of the characteristics of student's ability (Eraikhuemen & Ogumogu, 2014). Therefore, the higher the characteristics of the cognitive development stage, the more orderly and abstract the students' thinking.

The appropriate assessment to get information on student's thinking skill based on characteristics is by giving an appropriate test for measuring the thinking competence level. However, the current development of assessment is only based on the Classical Theory assumption in which scoring is performed step by step and student's score per item is gained

by adding the student's score in every step, and the skill is estimated by raw scores. Thus, an assessment which can cover the thinking skill level such as problem identification to assessment should be developed (Gok, 2010). Therefore, a physics problem-solving skill test instrument was developed for grade X students of high school. The purpose of the study was to produce an instrument to measure physics PSS in grade X students in their even semester and to get the characteristics of the physics PSS assessment instrument.

Method

This study is a developmental study with quantitative approach. The instrument development used in this study was the modified Orindo and Antonio model (Oriondo & Dallo-Antonio, 1998). The developed assessment instrument was a physics PSS test for grade X students in their even semester of 2016/2017 academic year.

Population

The study was performed in public high schools (or *Sekolah Menengah Atas Negeri* – *SMA Negeri*) in Kulonprogo Regency, Yogyakarta, i.e. *SMA Negeri* 1 Wates, *SMA Negeri* 2 Wates and *SMA Negeri* 1 Pengasih. The trial subjects were 281 students. The sample consisted of the students who had received similar tested materials in the three schools and they were selected not based on ranking.

The valid instrument was used in the form of a PSS test instrument packed in two packages of materials, each containing 30 questions with 8 anchor items of multiple-choice type reasonably ready for use in empirical testing. Testing is done by testing the instrument to 281 students.

The respondents were chosen from the class which had studied the materials of elasticity, static fluid, temperature and heat and optical equipment. They were classes X of *SMA Negeri* 2 Wates, SMA Negeri 1 Wates, and *SMA Negeri* 1 Pengasih Kulon Progo. The test results were analyzed by reference of the test using the criteria of acceptance of instrument suitability with Rash model, seen from the mean value of INFIT MNSQ (Mean Of Square) which ranged from 0.77 to 1.33 (Adams & Khoo, 1996, p. 30).

The trial sample in the analysis by IRT consisted of 281 students, who were required in IRT model research. Some experts consider that the bigger the sample size, the better the measurement result will be. One of the bases for using 281 students as the trial sample was Shin, who was using 200 to 1000 (Shin, 2009). Therefore, the 281 students used in this measurement was considered adequate.

Data Collection Technique

The instrument development was based on the aspects and sub-aspects of PSS test, including: (a) test design, (b) test trial, and (c) test revision and preparation. Meanwhile, the instrument designing stage consisted of: (1) needs analysis, (2) mapping, (3) drawing conclusion, (4) determining test purpose, (5) determining tested competencies, (6) determining tested materials, (7) preparing test answers, (8) writing items, (9) validating content by expert, (10) improving and preparing test, and (11) preparing scoring guide with Partial Credit Model (PCM). The trial stage consisted of: (1) determining trial subjects, (2) performing trial, and also (3) analyzing the trial result data based on IRT.

Figure 1 shows the test development stage. The test developed was a physical test used in high school with problem-solving aspect. The test was developed in the form of a multiple choice item consisting of 60 items including 8 anchor items. The test developed yielded 2 sets of problems with package A of 30 questions and package B of 30 questions. Each package has 8 anchor items.

The data analysis employed in this study was Partial Credit Model 1 PL (PCM 1-PL) for the testing item fitness of the physics PSS test for grade X students of high school. Based on IRT, the sample was adequate and good according to PCM 1-PL model (Adams & Khoo, 1996). The content validity analysis was performed qualitatively by material experts using Aiken index. The content validity analysis was performed qualitatively by material experts using Aiken's V index. Based on the index, the item was valid if the minimal Aiken's V is 0.87 (Aiken, 1980).

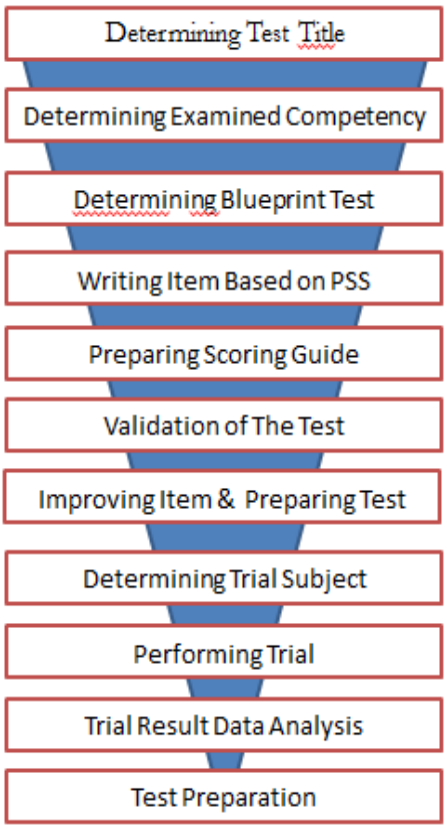


Figure 1. Phases of test development

The data analysis was performed on several aspects, including (1) the fitness of instrument items, (2) the reliability, (3) the item characteristic curve (ICC), (4) the difficulty index, and also (5) the total information function and standard error measurement (SEM). The goodness of the fit test for the overall test and testees (case/person) was based on the average INFIT Mean of Square (Mean INFIT MNSQ) and its standard deviation, or by the observation of the average INFIT *t* (Mean INFIT *t*) and its standard deviation. If the average INFIT MNSQ was approximately 1.0 and its standard deviation was 0.0 or the average INFIT *t* was approaching 0.0 and its standard deviation was 1.0, then the whole test fits the model. An item or testee/case/person fits a model in the INFIT MNSQ ranging from 0.77 to 1.30. An item was good if the difficulty index was over -2.0 or less than 2.0 (Hambleton & Swaminathan, 1985). The test reliability was tested by testing the information function and the following criteria presented in Table 2.

Table 2. Criteria of ideal score

Score Criteria Realibility	Category
>0.94	Excellent
0.91 – 0.94	Very Good
0.81 – 0.90	Good
0.67 – 0.80	Acceptable
<0.67	Questionable

Findings and Discussion

The development resulted in a problem-solving skill test with two sets of problems, coded A and B, each consisting of the materials of: elasticity, static fluid, temperature and heat, and also optical instruments. Table 3 shows the item distribution with eight items as the anchor items with the aspects of identification, planning, application, and also assessment.

Table 3. Distribution test

Subject		Elasticity	Static Fluid	Temperature and Heat	Optic
Aspect/ Sub aspect					
Identify	Distinguish	1a* 1b*	8a 8b	17a 17b	24a 24b
	Identify	2a 2b			25a *25b*
Plan	Formulate	3a 3b	9a 9b, 10a 10b	19a 19b	
	Devise	4a 4b			26a 26b
Apply and Execute	Connect	5a 5b	12a 12b, 11a 11b, 16a *16b*		28a 28b
	Apply		13a 13b	21a *21b*	29a 29b
				20a, 20b, 18a 18b	
	Analyze	6a 6b	14a* 14b*	23a* 23b*	27a*, 27b*
Evaluation	Investigate	7a *7b*		22a 22b	30a 30b
	Conclude		15a 15b		

The research product was validated by two assessment experts and five practitioners to assess the feasibility. Aiken index is in the range of 0.8 to 1.00. It can be interpreted that all of the items have good content validity and have supported overall content validity.

The fit goodness was tested for overall test items. The fitness of the overall test items used the principle developed by Adams and Khoo (1996, p. 30) based on INFIT Mean of Square (Mean INFITMNSQ) and its standard deviation or observing the average INFIT *t* (Mean INFIT t) and its standard deviation.

If the average INFITMNSQ was approximately 1.0 and its standard deviation 0.0 or the average INFIT *t* approached 0.0 and its standard deviation 1.0, the overall test fits PCM 1-PL model. Table 4 shows the average INFITMNSQ is 1.00 and its standard deviation 0.02, so the overall test fits PCM 1 PL model.

The fitness determination of each item followed the principle of Adams and Khoo (1996, p. 30) in which an item fits the model if INFIT MNSQ ranges from 0.77 to 1.30. With INFIT MNSQ as the item acceptance limit or fit according to the model (ranging from 0.77 to 1.30) and by using the INFIT *t* from -2.0 to 2.0, the items which met the goodness of fit were found. The INFIT MNSQ value ranged from 0.99 to 1.03. With INFIT MNSQ as the item acceptance limit or fit according to the model (ranging from 0.77 to 1.30), all of the 52 items fit the PCM.

Table 4. Testing the statistic fit parameter level

No	Test Parameter	Item estimation	Case Estimation
1	Average and std.deviation	-0.25 ± 0.28	0.17 ± 0.02
2	INFIT MNSQ	1.00 ± 0.02	1.00 ± 0.12
3	Outfit MNSQ	1.00 ± 0.02	1.00 ± 0.12
4	INFIT ZSTD	0.09 ± 0.75	0.06 ± 1.84
5	Average difficulty	1.00 ± 0.95	
6	Estimate Reliability	0.8	

The result of the reliability testing shows that the value of the reliability of the instrument is 0.28. Based on the relative value, the whole item is reliable as it corresponds to the reliability of the interpretation data of the Rasch model sufficiently categorized.

Figure 2 shows the goodness of item with an analysis by quest. Based on results of the analysis, it can be concluded that the entire test items are in accordance with the PCM model with the whole item being within the range of INFIT MNSQ PSS from 0.77 to 1.33 and using INFIT t with the limit of -2.0 to 2.0 in accordance with Figure 2 that no item exceeds the acceptance limit. In conclusion, 52 items fit the PCM model.

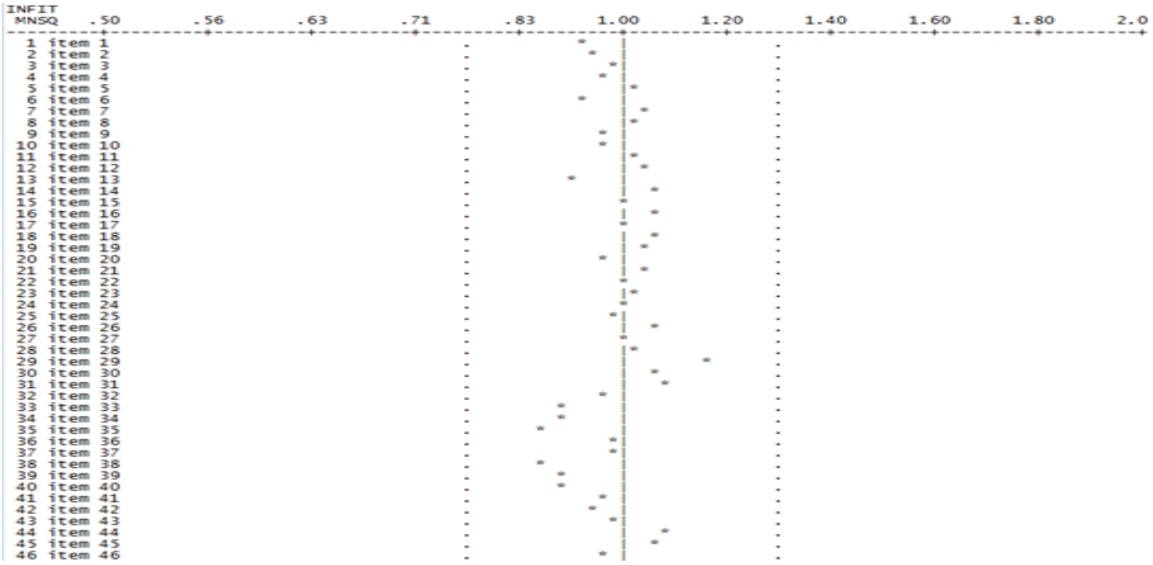


Figure 2. Goodness of fit instrument

Based on the result of analysis, the reliability of the instrument is 0.80. The reliability is adequate. The instrument has adequate strength and reliability because it consists of the items which have high information function (Hambleton & Swaminathan, 1985, p. 94). It may be because the test fits the skill of the tested students.

An item is categorized as good if the difficulty index is higher than -2.0 or less than 2.0 (Hambleton & Swaminathan, 1985, p. 36). Based on the analysis result, the items difficulty is between -0.95 and 1.0 with an average of 0 and standard deviation of 0.32. Therefore, based on the difficulty level, 52 items are good. The average difficulty of the aspect of problem-solving skills are shown by Table 5.

Table 5. Average difficulty of the aspect of problem-solving skills

Aspect	Difficulty
Identify	-0.13
Plan	-0.16
Apply and Execute	0.20
Evaluate	0.54

Construct validity is empirically proven by goodness of fit in the partial credit model (PCM). Table 4 shows the average value and standard deviation of INFIT MNSQ are 1.00 and 0.02, respectively, so the test fits PCM 1 PL. This means that the test is empirically valid. The test contains valid aspects of the PSS. This is because: (1) the items were developed consistently with the appropriate instrument item development procedure, (2) the items were developed from indicators derived from the aspects of the problem-solving skill and physics materials, (3) the test consisted of 52 items whose content validity was examined through expert judgment, and (4) the tryout respondents (students) worked on the test seriously (Istiyono, Mardapi, & Suparno, 2014). The difficulty level b for good item varies between -2.00 and 2.00. An item with the difficulty level of -2.00 is very easy, while that with the difficulty level of 2.00 is very difficult. Based on the test characteristics, the problem-solving skill test had the reliability coefficient, test information function, and estimation parameter which were reliable and had high stability.

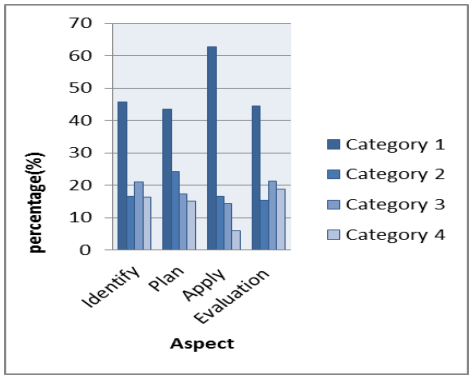


Figure 3. The percentage of difficulty level of aspect

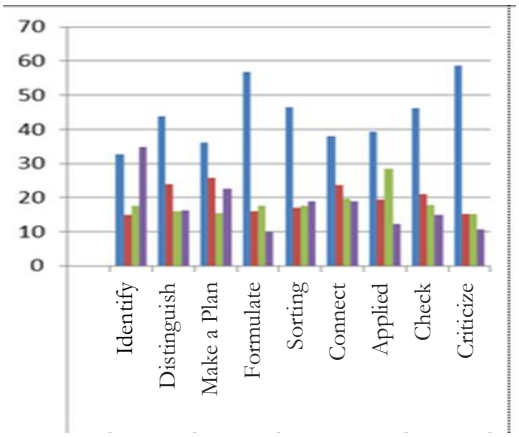


Figure 4. The percentage of difficulty level of sub-aspect

Figures 3 and 4 show the percentage of the aspects and sub-aspects that have been tested. The percentage of the results indicates that the frequency of students' responses to the per item categories of each aspect and the sub-category is put into category one, two, three and four. The first category states that the frequent answers are with a score of one whereas a score of four is expressed by the fourth category.

The percentage of each difficulty level of each item is shown in Figure 3. It shows that the highest difficulty level is in the application aspect. Category 1 percentage shows that most students answer correctly in score 1, so the item is difficult. Figure 3 shows that the percentage of the application in category 1 is 64 and that in category 4 is 6. Figure 4 shows the level of difficulty of each aspect of the problem-solving skill.

The differences between the classical theory and the modern theory in educational assessment can be illustrated by five students

A, B, C, D, and E taking the test as many as 5 items with five alternatives type. The wrong item was given a score of 0 and a maximum of four is given to the correct answer.

The most difficult aspect is the evaluation and implementation aspect. This shows that the students' problem-solving skill in evaluation and implementation aspects is still low.

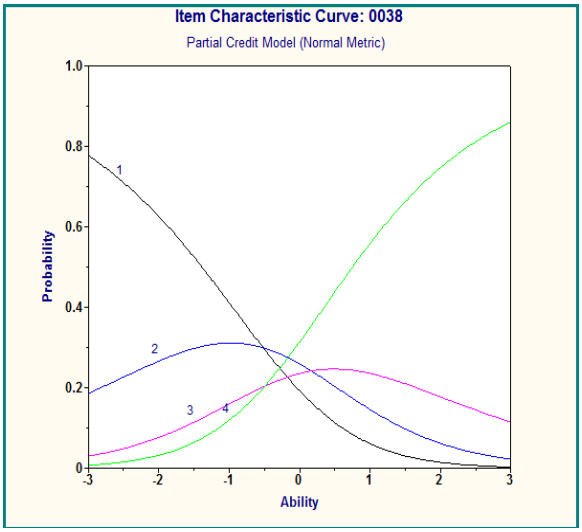


Figure 5. ICC of item no. 38

The characteristic of the item is indicated by the item characteristic curve (ICC) and the difficulty index. Based on the result of the ICC analysis, 52 items are equivalent to the number of the questionnaire items developed. Figure 5 shows an example of ICC for item 38. It shows that in Category 1, the ability of most of the students is very low (θ = -3), in Category 2, the ability of most of the students is low (θ = -1), in Category 3, the ability of some students is high (θ = 1), in Category 4, the ability of most of the students is very high (θ = 3). The difficulty level ranges from small to large sequential categories 1, 2, 3, and 4.

Based on Figure 6, the measurement information is in the range of the ability of -1.3 to 2.7. Therefore, the test instrument is suitable to be used for the students with -1.3 to 2.7 so that in that range, information function shows the ability level estimated by the test (Thorpe et al., 2007, p. 179). The assessment of learning achievement in physics is an assessment of the results of the physics learning process which is a number that describes the characteristics of individual students.

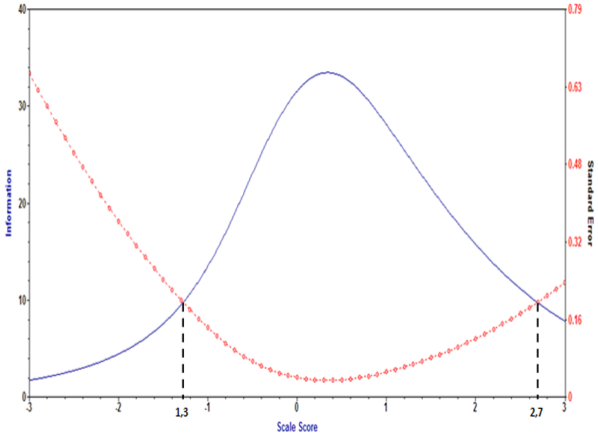


Figure 6. Information function & Standard Error Measurement (SEM)

The relationship between the information function and SEM shows the grand contribution of the test to expressing the latent ability as measured by the test. The greater the value of IF given by the item on the test, the fewer the measurement errors. Therefore, the test is suitable to be used in measuring students' problem-solving skill in the ability categories of medium, low, and high.

Based on the discussion, the test is feasible to use in measuring students' PSS, because: (1) the developed items were consistent with the appropriate instrument item development procedure, (2) the items were developed from problem-solving indicators, (3) the test consists of 52 items whose content validity was examined through expert judgment, and (4) the tested respondents (students) did the test seriously because they were observed by their teachers. This was consistent with the finding of Istiyono et al. (2014). Therefore, the instrument is expected to be able to be used to measure problem-solving skill appropriately. Problem-solving assessment can help students understand a problem quickly (Gok, 2010). Thus, this instrument can be used to measure the exact problem-solving skills.

Conclusion

The problem-solving skill instrument developed in the form of a multiple choice test is based on the problem-solving skills in the physics materials of elasticity, static fluid, temperature and heat and optics consisting of set A and set B each wih 8 anchor items has 52 items.

The problem-solving skill test fulfills the content validity by expert judgment and has empirical evidence of construct validity which fits Partial Credit Model (PCM) based on polytomous data of four categories. The reliability PSS test has met the requirement (reliability coefficient of 0.79). In terms of difficulty level of 52 test items, it is good, between -2 and +2. Thus the test is suitable for measuring the problem-solving ability of students in medium, low and high category of tray.

Based on the information function, the PSS test is appropriate for measuring students' problem-solving skill from -1.3 to 2.7 with a good item difficulty level. Therefore, the test is qualified and so it can be used to measure the physics problem-solving skill of grade X students of high school.

References

Adams, R. J., & Khoo, S.-T. (1996). *Quest: The interactive test analysis system version 2.1*. Victoria: Australian Council for Educational Research.

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria: ASCD.

Cangelosi, J. (1995). *Merancang tes untuk menilai prestasi siswa*. (D. Tedjasudhana, Ed.). Bandung: Institut Teknologi Bandung.

Carvalho, C., Fíuza, E., Conboy, J., Fonseca, J., Santos, J., Gama, A. P., & Salema, M. H. (2015). Critical thinking, real life problems and feedback in the sciences classroom. *Journal of Turkish Science Education*, *12*(2), 21–31.

Eraikhuemen, L., & Ogumogu, A. E. (2014). An assessment of secondary school physics teachers conceptual understanding of force and motion in Edo South Senatorial District. *Academic Research International*, *5*(1), 253–262.

Gok, T. (2010). The general assessment of problem solving processes and metacognition in physics education. *Eurasian Journal of Physics and Chemistry Education*, *2*(2), 110–122. Retrieved from http://www.eurasianjournals.com/index.php/ejpce

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

Helaiya, S. (2010). *Development and implementation of life skills programme for student teachers*. Vadodara: Maharaja Sayaji Rao University of Baroda.

Istiyono, E. (2016). The application of GPCM on MMC test as a fair alternative assessment model in physics learning. In *Proceeding of the 3rd International Conference on Research, Implementation and Education of Mathematics and Science (ICRIEMS), 16-17 May 2017* (pp. 25–30). Yogyakarta: Universitas Negeri Yogyakarta. Retrieved from http://seminar.uny.ac.id/icriems/sites/seminar.uny.ac.id.icriems/files/prosiding/PE-04.pdf

Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (Pys-THOTS) peserta didik SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 1–12. https://doi.org/10.21831/pep.v18i1.2120

Nadapdap, A. T. Y., & Lede, Y. (2016). Authentic assessment of problem solving and critical thinking skill for improvement in learning physics. In *Proceeding of International Seminar on Science Education (ISSE), 29 October 2016* (pp. 37–42). Yogyakarta: Universitas Negeri Yogyakarta.

Oriondo, L. L., & Dallo-Antonio, E. M. (1998). *Evaluation educational outcomes*. Manila: Rex Printing Compagny.

Piaget, J. (2005). *The psychology of intelligence* (Electronic version). Taylor & Francis.

Pólya, G. (1957). *How to solve it: A new aspect of mathematical method*. Doubleday: Garden City.

Regulation of Minister of Education and Culture No. 59 of 2014 on the curriculum 2013 of senior high school/Madrasah Aliyah (2014). Republic of Indonesia.

Shin, S.-H. (2009). How to treat omitted responses in Rasch model-based equating. *Practical Assessment, Research & Evaluation*, *14*(1), 1–8. Retrieved from http://pareonline.net/getvn.asp?v=14&n=1

Slameto. (2010). *Belajar dan faktor-faktor yang mempengaruhi*. Jakarta: Rineka Cipta.

Suryabrata, S. (2002). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi Offset.

Thorpe, G. L., McMillan, E., Sigmon, S. T., Owings, L. R., Dawson, R., & Bouman, P. (2007). Latent trait modeling with the Common Beliefs Survey III: Using item response theory to evaluate an irrational beliefs inventory. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *25*(3), 175–189. https://doi.org/10.1007/s10942-006-0039-9

Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, *87*(2), 320–32.

REiD (Research and Evaluation in Education), 3(2), 2017, 124-132
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

The implementation of population education in senior high school

***[1]Claver Nzobonimpa; [2]Zamroni**
*Department of English Language and Literature, Faculty of Languages and Social Sciences,
Université du Burundi (National University of Burundi)
UNESCO Avenue No. 2, P.O. Box 1550 Bujumbura, Burundi
*Email: nzobonimpacl@yahoo.fr

Submitted: 14 July 2017 | Revised: 27 December 2017 | Accepted: 06 February 2018

Abstract

This research aimed to evaluate the implementation of Population Education in senior high school in terms of (1) learning process, (2) learning materials, (3) evaluation process, (4) course outcome, (5) teachers' role, (6) perception of Population Education, and (7) factors supporting and inhabitting Population Education. The research subjects were one teachers' supervisor, three teachers, and 65 students. The data were collected through questionnaires, interviews, and documentation and analyzed quantitatively using descriptive statistics. The qualitative data collected through interviews were used for deeper explanation. The research findings were: (1) the teaching process was not quite appropriate, (2) materials for Population Education were available and efficient, (3) the evaluation process was not appropriate, (4) the students were satisfied with the teachers' role, (5) the students' perception of Population Education was very positive, and (6) the constraints in Population Education included (a) limitation in time, (b) too many extracurricular activities, (c) rapid change of data, and (d) the validity of materials.

Keywords: *population education, implementation, learning process, integration*

How to cite item:

Nzobonimpa, C., & Zamroni, Z. (2017). The implementation of population education in senior high school. *REiD (Research and Evaluation in Education), 3*(2), 124-132. doi:http://dx.doi.org/10.21831/reid.v3i2.10024

Introduction

Education is very important for human being. Moore (2015, p. 1) says that: '*Changes in society are often in more demands being placed on our education system'*. Further, as stated in Law No. 20 of 2003, Indonesian national educational system ensures equal opportunity, improvement of quality, relevance and efficiency in education to meet various challenges in the development of local, national, and global lives changes (UNESCO, 2015, p. 1). Syamsudin, Budiyono, and Sutrisno (2016, p. 26) inform that the goal of education in Indonesia is to develop learners' potentials so that they become Indonesian individuals with faith and fear of God, noble morals, good health, great knowledge, high competency, creativity, and independence, and become individuals who are democratic and responsible. In order to reach its education objectives, Indonesian government elaborates the curriculum which contains the objectives and strategy to achieve the education goals. In line with this opinion, Indonesian National Education Law of 2003 defines curriculum as "*...a set of plan and regulations about the aims, content, materials of lessons and the methods employed as guidelines for implementation of learning activities to achieve given education objectives"* (Dharma, 2008).

Population Education is one of the teaching programs delivered in schools. It is a

program introduced due to the rapid population growth in both the industrial and developing countries. In early 1960s, the study of human reproduction, birth control, and also investigation of the cause and effect of population was included into the school curriculum (Sulistyo, 1997, p. 26). In secondary schools, the government integrated Population Education topic into six subjects: Biology, Geography, Economics, Civics, Physical Education, and Anthropology with the use of the integrative approach.

Though Population Education has been introduced in Indonesian formal education many decades ago, some problems still occur. There are still significant differences between the ideal situation (self-reported) and the actual practice related to teachers' roles in teaching Population Education. This difference indicates that there are role conflicts for teachers in teaching Population Education. Some observable barriers related to the implementation of Population Education are the lack of teachers' knowledge and skill, and also the lack of teachers' autonomy in carrying out teaching activities.

Education

John Dewey (Ornstein & Levine, 1989, p. 10) considers education as a social process by which the groups of immature members, especially children, learn to participate in a group life. Thus, through education, children receive knowledge about their cultural heritage and learn to use it in problem solving. Hills (1986, p. 50) says that education has two principles: passing on knowledge from one generation to the next, and providing the people with skills which enable them to analyze, diagnose, and question something. Education, in the narrowed sense, is regarded to be equivalent to instruction. It consists of 'specific influences' given consciously to bring in the development and growth of the students. In general, education aims to transmit a common set of beliefs, values, norms, and understanding from the adult to the youth. Morality, on the other hand, aims to maintain the order in a society; to respect people as well as regard them holistically (Nayef, Yaacob, & Ismail, 2013, p. 165).

Population Education

Viederman (V. K. Rao, 2001, p. 31) says that Population Education may be defined as an educational process which assists persons to (a) learn causes and consequences of population problems; (b) define the nature of the problems associated with population process and characteristics; and (c) assess the positive and effective means by which the society as a whole and he/she as an individual can respond to the areas that influence these processes in order to enhance the quality of life. Rao (2004, p. 34) says that: '*Population Education is an educational program which provides for a study of population situation in the family, community with the purpose of developing in the students' rational and responsible attitudes and behavior toward that situation'*. Based on the definition, we can understand that Population Education is a program which provides a study of population situation at various levels. It also intends to develop rational and responsible attitudes and behavior to that situation.

Learning

Learning is identified as some kinds of change in behavior which is relatively long lasting. According to Schunk (2012, p. 3), the definition of learning is '*an enduring change in behavior, or in the capacity to behave in a given fashion, which result from practice or other form of experience'*. Learning aims at changing the behavior of the learner. Learning is the main activity organized in school which has three main criteria: (1) learning involves change, (2) learning endures over times, and (3) learning occurs through experience.

Illeris (2009, p. 14) distinguishes the definition of learning into four. First, learning can refer to the results of individual learning processes. Second, learning refers to individual psychological processes that lead to alterations or results described as meaning. Third, learning, as well as processes of learning, refers to the interaction process among individuals, his/her material, and social environment described as meaning. Fourth, learning and process of learning are used identically with the word teaching. It may be interpreted as a result of tacit short circuit between what is taught and what is learned.

In the discussion of learning activities, Assan (2014, p. 340) insists that learning activities, especially in adults, have three features, including the facts that: (1) the learners develop different outlooks and approaches with maturity and/or experience; (2) the learners reveal different degrees of independence in their learning; (3) the learners exhibit a different amount of involvement in, or different approaches to, learning tasks. The type of involvement is often dependent upon the context in which the learning activity takes place.

As far as learning theories are concerned, we distinguish the following learning theories:

Self-directed learning. Borich (2000, p. 273) has defined self-directed learning as an approach to teaching and learning that actively engages students in the learning process to acquire the high levels of behavioral complexity outcome. Mohammadi and Araghi (2013, p. 75) assert that self-directed learning refers to any self-teaching projects in which the learner establishes his specific goal, decides how to achieve it, finds the relevant resources, plans his strategies, and maintains his motivation to learn independently. Bear (2012, p. 28) argues that self-directed learning is a process which occurs when individuals take initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and also evaluating learning outcomes.

Cooperative learning. Unlike self-directed learning, cooperative learning is defined as activities that involve groups of students jointly working through assigned tasks (after receiving instruction from the teacher) until all of the group members have successfully mastered and completed them (Johnson, et al., in Thanh, 2014, p. 3).

Discovery Learning. Joy (2014, p. 32) explains that learning happens by discovering, which prioritizes reflection, thinking, experimenting, and exploring. He also suggests that the discovery learning approach is closer to the concepts of exploration, discovering, invention and the 'knowledge cannot be transferred from one person to another' concept; instead, a student needs to experience an event in order to make it truly meaningful.

Perception

In the perspectives of social psychology, Walgito (2010, p. 99) defines perception as the process of organizing, and interpreting the stimulus received into something meaningful. In perception, the stimulus may come from the outside of the individuals (external) or within the individuals (internal). Furthermore, Mozkowit and Orgel (Walgito, 2010, p. 101) argue that perception is a global response to a stimulus. From those definitions, perception is viewed as the response to a stimulus or surroundings. Then these responses will be interpreted as meaningful information related to the stimuli.

Teacher's Role in Population Education

Malik, Murtaza, and Khan (2011, p. 784) determine the teachers' role in learning-teaching processes as the persons who are responsible to ensure whether the teaching process puts emphasis on course context, interpersonal relationship, or on classroom discipline and control. The following cases are also taken into consideration by teachers: (1) the kind of learning being promoted by putting emphasis on the acquisition of skill, facts or understanding; (2) the pattern of communication in the classroom; and (3) students' communication, by keeping eye on the way in which educational tasks are organized.

Hudgins et al. (1983, p. 489) distinguish six roles of a teacher in the classroom. First, a teacher is a transmitter. In this role, his duty is to transmit factual information to students. Second, he is a socializer; he supervises the development of moral values and norms of his students. Third, he is an initiator and administrator of goals; he initiates and administers long-range and short-run activities and goals of the class membership. Fourth, he is an evaluator. He evaluates his students' academic performance. Fifth, he is a motivator; he motivates his students to realize their achievement potential. Sixth, he is a disciplinarian. His duty is to discipline and apply sanctions in response to the class members' behavior.

Method

The main aim of this research is to find out the implementation of Population Education in senior high school. This research used a mixed method (quantitative data were analyzed under descriptive statistics method, then supported by qualitative data analysis). The basic assumption is that the use of both quantitative and qualitative methods in combination may provide a better understanding of the research problem and question (Creswell, 2010).

The research was conducted in a senior high school in Yogyakarta Special Region, Indonesia, from January to April 2016. The sample consisted of 65 students of class XI, three teachers (Sociology teacher, Economics teacher, and Geography teacher), and also one supervisor.

Documentation was used to collect the Population Education curriculum, students' books, and teachers books. The sample of the research is presented in Table 1.

Table 1. Research sample

Data source	Rate
Students	65
Teachers	3
Principals	1
Total	69

Research Variables

In this research, the aspects evaluated are: (1) the efficiency of the learning/teaching materials, (2) the appropriateness of the learning/teaching process and evaluation process, (3) the teachers' and students' satisfaction on Population Education outcome, (4) the efficiency of the evaluation process, (5) Population Education outcome, (6) the students and teachers' appreciation of Population Education, (7) teachers' role, and (8) the factors that facilitate or inhibit the learning process of Population Education.

Data Collection Techniques

This research used a variety of data collection techniques, i.e. questionnaires, observation, and interview. In order to collect the quantitative data, questionnaires were given to 65 students. The qualitative data were collected through classroom observations and an interview with four teachers of Sociology, Geography, Economy and one teacher who is in charge of monitoring the social studies program.

Research Instruments

The research involved the following instruments. The first one is *observation guide and checklist*. Observations were conducted in the beginning of the semester. Through these observations, the researchers collected information about school and its Population Education program. The researchers also checked the teacher's materials, students' text books and some teachers' facilities through checklist.

The second instrument is *a questionnaire*. The students were given an open and closed questionnaire. The questions were related to (1) the efficiency of the learning material, (2) the appropriateness of learning/teaching process, (3) teachers' and students' satisfaction on Population Education outcome, (4) efficiency of evaluation, (5) Population Education outcome, (6) students' and teachers' appreciation of Population Education, (7) teachers' role, and (8) the factors that facilitate or inhibit Population Education learning processes. The third instrument is *an interview guide*. The topics of the interview were identical with the questionnaire evaluation aspect.

Validity and Reliability of Instruments

Validity assessment was required to provide an evidence related to whether the instrument indeed accomplishes what it is supposed to accomplish (Teo, 2013). In this research, the face validity and content validity were used to validate the instruments by involving two experts in Social Studies. In order to check whether the research instruments measured what it was supposed to measure, a tryout test was administered. The tryout results had allowed the researchers to revise the content and form of some variables.

Data Analysis Techniques

The questionnaire applying modified Likert scale which is proposed by Mardapi

(2008, p. 23) was administered to 65 students and analyzed using descriptive statistics. Table 2 shows the criteria for learning process, material, course, outcome and also perception of Population Education.

This analysis was followed by three key stages of analyzing qualitative data. Miles and Huberman in Irambona and Kumaidi (2015, p. 121) explain that the three stages of qualitative data analysis are data reduction, data display, and conclusion formulation. The qualitative data were reduced to make them simpler to analyze, then were summarized and formulated to a conclusion. This analysis was done during data collection, as well as after all of the data had been gathered.

Findings and Discussion

Findings

Population Education Learning Process

Based on the students' stand point, the learning process of Population Education is less appropriate. The mean score of the students' rating is 29.6, which means that most students chose 'sometimes' category. Based on the interview with Geography, Economics and Sociology teachers and the teachers' supervisor, it is discovered that Population Education is not planned as an integrated lesson. The researchers also discover that there are some opinions related to Population Education, including: (1) Population Education is not popular, (2) some teachers do not have any concern in teaching Population Education in their courses, (3) Population Education is not a prominent material in social science class, (4) Population Education course taught only concerns Indonesia and East Asia issues.

Learning Materials

Questions were asked to the students in order to discover the efficiency of learning material and sources which are used in Population Education learning process. It is revealed that students are satisfied with the materials. This is reflected by the number of students who chose 'always' and 'often'. There are 43.07% of the students who chose 'always' category. Meanwhile, 24 students or 36.92% of all students chose 'often' category. The rest of the sample is in the two remaining categories. There are 12 students choosing 'sometimes' and only one student chose 'never' category. The mean score of learning material efficiency is 14.17 and it is included in 'often category'.

According to the interview with the teachers, the researchers discovered that the material/books related to Population Education are easily found. It is also discovered that mass media help teachers to improve and update their learning material. Television, newspapers, internet and other information technology help the teachers and students as the learning sources of references.

Evaluation

The objective of this research is to find out whether the students are given assignments and instruction to discuss population issues inside/outside of the class. The researchers found that the students' opinion on the evaluation appropriateness is less appropriate. There are only seven students (10.77%) who chose 'always' category. Meanwhile, 16 out of 65 students (24.61%) chose 'often' category, while more than half of the students chose 'sometimes' and 'never'.

Table 2. The criteria of learning process, material, course, outcome and perception

Score X	Categories	Predicate
X≥M+1SD	Strongly agree/ Always	(Very)Satisfying/Positive/Good/Appropriate
M≤X<+1.5 SD	Agree/Often	Satisfying/Positive/Good /Appropriate
M-1.SD≤X<M	Disagree/Sometimes	(Less) satisfying/appropriate
		Negative/Bad
X<M-1.SD	Strongly disagree/Never	Not satisfying/ Not appropriate/Very Negative/Very bad

Based on the interview, the teachers gave assignments related to Population Education issues, but only if the subjects being taught (Geography, Sociology, and Economics) contained population issues. It is also discovered that Population Education has a limited time allocation. As a consequence, the items introduced in assignments are quite few. The teachers plan a discussions in class by giving a specific topics. Mostly, the topics are given to be discussed in class. They are not encouraged to do discussion on Population Education outside the classroom.

Teachers' Role in Population Education

The role of teachers in Population Education is appropriate. In fact, the indices of appropriateness are close to the 'appropriate' category, i.e. 0.87. The students are satisfied with the role of teachers as managers, guiders, instructors, judges, and parents of Population Education learning process.

Perception on Population Education

It is discovered that the perception of Population Education's mean score is 27.38 which is in 'strongly agree' category. There are 31 out of 65 students (47.70%) who chose 'strongly agree'. The remaining students are in the category of 'agree'. It can be concluded that students are satisfied with the implementation of Population Education learning process. Based on the interview, it is discovered that students are interested in Population Education and the teachers expressed that population is interesting since it deals with actual issues which are faced by the nation which students meet in their daily life.

Factors Facilitating Population Education

There are several factors that facilitate Population Education in senior high school, including (1) students' motivation to learn Population Education, (2) materials availability in school library, (3) time allocation, and also (4) extracurricular activities. The factors which inhibit Population Education are limited time allocation, unfamiliarity of the teachers with the topics, and also invalid/out-of-date materials which are avaliable in the learning process.

Discussion

Learning Process in Population Education

In general, some Population Education topics are integrated in some courses (Geography, Economics, and Sociology). Though those three courses are integrated with Population Education, only Geography teachers prepare an entire lesson about Population Education because Population Education is not the main focus of the courses. Population Education is a small part of those courses with limited time allocation.

Materials in Population Education Learning

Based on the findings, materials in Population Education are available. Concerning students' opinion on Population Education materials related to its availability and usage, the mean score is in 'often' category with the score of 14.17. The sources for Population Education learning material are not only available in school library, but also are collected from the Internet and mass media

Evaluation in Population Education

Population Education is integrated in Geography, Economics, and also Sociology courses. From the findings, students' opinion related to the evaluation process is in the category of 'sometimes'. The examinations and assignments which are given in these courses contain few items about Population Education. The evaluation of Population Education shows that teachers included few Population Education items in their exams and assignment. It is also discovered that the time allocation for Population Education in class is insufficient.

Teachers' Role in Population Education

In this section, it has been found that students and teachers are satisfied with the role of teachers in Population Education and other courses. In fact, teachers have the roles of being managers, guides, instructors, judges, and parents. In this research, it has been realized that teachers give instructions to students at the beginning of each lesson. Teachers take strategies to help and guide students. In order to install motivation to students, they bring

hot news as the topics. This allows students to be motivated and participated more in the class. As parents, teachers care about the students' future life by bringing the topics that are concerned with their future to discuss, such as unemployment and awareness on competition in the work field.

However, it has been realized that students are not asked and reminded to make discussion about population issues out of the class. Discussions outside the class reinforce knowledge and help students to learn about Population Education in a natural way. This would extend their awareness of population issues that they may face in their life.

Perception of Population Education

The research found that students are interested in Population Education and find that it is very useful to learn it. Perception of Population Education is one of the variables that is very favorable (most students chose 'strongly agree'), which means that students' perception of Population Education is very positive. Population Education is connected to every aspect and issue of life. To be very clear, Population Education is very factual, so that it can be easy to understand. Teachers find that Population Education is very important to students. Through Population Education, the teachers have a chance to talk about the real life, and also to draw a picture of Indonesia in terms of population.

Factors Supporting and Inhibiting the Teaching of Population Education

There are some factors that support the teaching of Population Education. The first factor is that Population Education deals with everyday-life issues. Students get motivated to learn it since it talks about everyday life in simple and factual way. Second, the dynamism of Population Education topics can be found everywhere. It is not difficult to get information about Population Education topics as well as population issues. The third factor is the availability of materials to be used.

On the other hand, there are some factors which constrain to Population Education. The first factor is time allocation. The time allocation for Population Education is limited, so that the topics of Population Education are not suffieciently exploited. Second, the students are obliged to join various extracuricullar activities that spend their time, so that the time to revise of their learning material is limited. The extracurricular activities are paid by students themselves; there is no support from school. The students are asked to do what they can do to get the money to pay the cost for those activities. They are stressful with these extracurricular activities and it can diminish their motivation to learn. Third, the teachers are not familiar with Population Education. It has also been found that the teachers are unable to catch the dynamism of Population Education materials. Fourth, although some books of Population Education are available in the newest edition, the content of the books are relatively out-of-date, while the information which is gained from the mass media and the Internet are sometimes not valid.

Conclusion and Recommendations

Conclusion

Based on the findings and discussion, it can be concluded that: (1) the teaching process of Population Education in senior high school is less appropriate; (2) the Population Education materials are avaliable in the library, mass media, and the Internet; (3) the evaluation in Population Education learning is less appropriate. The items are included in the assignments and examinations, because the teachers allocated limited quota for Population Education items; (4) the teachers' role in Population Education is appropriate, especially in bringing the actual topics to motivate the students; (5) the students' and teachers' perception of Population Education is very positive (they consider Population Education an important thing to learn); (6) the favorable factors of Population Education are the recency of the topics, dynamism of the topics, and avaliability of the materials. The factors which inhibit Population Education are the limited time allocation, teachers' unfamiliarity with the topics, and invalid/out-of-date materials.

Recommendations

Referring to the research results, some recommendations are proposed, including: (1) Additional time allocation is needed for Population Education; (2) students and teachers should get more time to be involved in population issues through seminars; (3) the government should support the Population Education teachers with up-to-date data in order to ensure the validity; (4) the authorities need to reduce the time allocation for extracurricular activities; (5) teachers should be more consistent in their teaching activities by inviting students to be more concerned with population issues.

References

Assan, T. B. (2014). Perceptions of lecturers on quality assurance in higher education teaching and learning process. *International Journal of Educational Sciences*, *7*(2), 339–347. https://doi.org/10.1080/09751122.2014.11890196

Bear, A. A. G. (2012). Technology, learning, and individual differences. *Journal of Adult Education*, *41*(2), 27–42.

Borich, G. D. (2000). *Effective teaching methods*. London: Prentice Hall.

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

Dharma, A. (2008). Indonesian basic education curriculum: Current content and reform. *Presented in Roundtable Discussion in Retrac Governing Board Meeting at Institut Aminuddin Baki, Genting Highland, Malay-sia, on 27 August 2008*. Jakarta: Ministry of National Education. Retrieved from http://www.ibe.unesco.org/curricula/indonesia/io_befw_2008_eng.pdf.

Hills, P. J. (1986). *A dictionary of education*. New York, NY: Routledge & Kegan Paul.

Hudgins, B. B., Phye, G. D., Schau, C. G., Theisen, G. L., Ames, C., & Ames, R. (1983). *Educational psychology*. Itasca, IL: Peacock Publishers.

Illeris, K. (2009). *The three dimensions of learning: Contemporary learning theory in the tension field between the cognitive, the emotional and the social*. Malabar, FL: Roskilde.

Irambona, A., & Kumaidi, K. (2015). The effectiveness of English teaching program in senior high school: A case study. *REiD (Research and Evaluation in Education)*, *1*(2), 114–128. https://doi.org/10.21831/reid.v1i2.6666.

Joy, A. (2014). Impact of discovery-based learning method on senior secondary school physics. *IOSR Journal of Research & Method in Education*, *4*(3), 32–36. Retrieved from www.iosrjournals.org.

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

Malik, M. A., Murtaza, A., & Khan, A. M. (2011). Role of teachers in managing teaching learning situation. *Interdisciplinary Journal of Contemporary Research in Business*, *3*(5), 783–833. Retrieved from http://journal-archieves8.webs.com/783-833.pdf.

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

Mohammadi, P., & Araghi, S. M. (2013). The relationship between learners' self-directed learning readiness and their English for specific purposes course accomplishment at distance education in Iran. *Studies in Self-Access Learning Journal*, *4*(2), 73–84. Retrieved from http://sisaljournal.org.

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

Nayef, E. G., Yaacob, N. R. N., & Ismail, H. N. (2013). Taxonomies of educational objective domain. *International Journal of Academic Research in Business and Social Sciences*, *3*(9), 165–175. https://doi.org/10.6007/IJARBSS/v3-i9/199.

Ornstein, A. C., & Levine, D. U. (1989). *Foundations of education*. Dallas, TX: Houghton Mifflin.

Rao, D. B. (2004). *Teachers' population education awareness*. New Delhi: Discovery Publishing House.

Rao, V. K. (2001). *Population education*. New Delhi: A.P.H. Publishing Corporation.

Schunk, D. H. (2012). *Learning theories: An educational perspective*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Sulistyo, D. (1997). *Role perception and professional commitment of high school population education teachers: A case study in Yogyakarta Province, Indonesia*. Doctoral dissertation, Florida State University, Tallahassee, FL.

Syamsudin, A., Budiyono, B., & Sutrisno, S. (2016). Model of affective assessment of primary school students. *REiD (Research and Evaluation in Education)*, *2*(1), 25–41. https://doi.org/10.21831/reid.v2i1.8307

Teo, T. (2013). *Handbook of quantitative methods for educational research*. Dordrecht: Sense Publishers.

Thanh, P. T. H. (2014). *Implementing cross-culture pedagogies: Cooperative learning at Confucian heritage cultures*. Dordrecht: Springer Science Business Media.

UNESCO. (2015). *Transforming teaching and learning in Asia and the Pacific: Case studies from seven countries*. (E. Hau-Fai & U. Miura, Eds.). Bangkok: The United Nations Educational, Scientific and Cultural Organization and UNESCO Bangkok Office. Retrieved from http://unesdoc.unesco.org/images/0023/002329/232909E.pdf

Walgito, B. (2010). *Pengantar psikologi umum*. Yogyakarta: Andi.

REiD (Research and Evaluation in Education), 3(2), 2017, 133-143
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Discrepancies in assessing undergraduates' pragmatics learning

Oscar Ndayizeye
Higher Teacher-Training School of Burundi, (Ecole Normale Supérieure (ENS) du Burundi)
Boulevard du 28Novembre, B.P. 6983 Bujumbura, Burundi
Email: ndaosca@yahoo.fr

Submitted: 15 June 2017 | Revised: 28 December 2017 | Accepted: 03 January 2018

Abstract

The purpose of this research was to reveal the level of implementation of authentic assessment in the pragmatics course at the English Education Department of a university. Discrepancy Evaluation Model (DEM) was used. The instruments were questionnaire, documentation, and observation. The result of the research shows that respectively, the effectiveness of definition, installation, process, and production stages in logits are -0.06, -0.14, 0.45, and 0.02 on its aspect of the assessment methods' effectiveness in uncovering students' ability. Such values indicate that the level of implementation fell respectively into 'very high','high', 'low', and 'very low' categories. The students' success rate is in 'very high' category with the average score of 3.22. However, the overall implementation of the authentic assessment fell into a 'low' category with the average score of 0.06. Discrepancies leading to such a low implementation are the unavailability of the assessment scheme, that of scoring rubric, minimal (only 54.54%) diversification of assessment methods, infrequency of the lecturer's feedback on the students' academic achievement, and the non-use of portfolio assessment.

Keywords: *authentic assessment, program evaluation, pragmatics, Rasch model*

How to cite item:

Ndayizeye, O. (2017). Discrepancies in assessing undergraduates' pragmatics learning. *REiD (Research and Evaluation in Education), 3*(2), 133-143. doi:http://dx.doi.org/10.21831/reid.v3i2.14487

Introduction

Writing, for some people, springs out from something else, and the motivation to write this article is remote to 2014 when the authors audited a pragmatics course in English Language and Literature Study Program, Faculty of Languages and Arts of a university. During that time, they observed many but a thing among which the use of (a) classification by (Yule, 1996, pp. 47–48); (b) students' classroom presentations, during which each student was given a sheet used to comment on the presenters' content clarity and the language use in general, and after presentations, students were given a chance to comment/ read aloud their reflections on the previous presentations; (c) a detailed syllabus downloadable from the university's e-data of the staff, giving details on the assessment schema in that course whose assessment comprised students' attendance, class participation, assignments, mid-semester exam (which actually was a take-home exam), and final exam; and (d) a course book written by Yule (1996), entitled *Pragmatics*.

As the authors remarked, the characteristics previously featured are those indicating the authentic assessment of Yusuf (2015, pp. 292–293). However, with this pre-survey insights, Yusuf could not tell whether what he observed was really an authentic assessment being implemented in a pragmatics course. In 2017, wishing to discover more about the au-

thentic assessment as the authors observed that such assessment was quasi-absent in the assessment of linguistics-related course in the first author's country, they decided to go back to the Faculty of Languages and Arts, especially in the 5th semester in which pragmatics course was administered in the English Language and Literature Study Program of the university to investigate the issue.

In (higher) education, the solutions to assessment-related problems can be investigated in a series of aspects, such as, how lecturers may track plagiarism in students' assessment tasks, the development of fair assessment criteria/rubrics, the implementation of authentic assessment, and the impact of students' right to sue educators to the court and how this impedes on assessment. The list of these assessment-related perplexing issues in Indonesian (higher) education system or in the first author's country of origin is far from being exhaustive.

Assessment is a process that is integral part of the logic in which the lecturers' and their students' roles are to be played maximally for the learning to take place. The normal flow is that the lecturers give assessment tasks, and the students do them, and ideally this flow goes on until the students graduate. The problem arises when the two main parties in the teaching-learning process have different perception of some issues.

For example, the views on assessment sometimes diverge as lecturers might view it as a motivation for learning, while their students might see it as the emptiness of any motivation to improve learning but that it is only marking-grounded; and this has also become Fry, Ketteridge, and Marshall's (2009, p. 133) observation. Even among assessors, divergence does also exist. One trend of academics still thrive to use tests (exams) where students give short-answers while another advocates for real-life assignments that result in students' competency, knowledge and interest building. The academics in the last group even label short-answer exams as the traditional practice of assessment. Real-life assignment advocates also stress how this type of activities is related to motivating learning via well-timed and consistent feedback.

Whichever views, it is urgent to see the role of authentic assessment in language classes and how feedback might enhance learning improvement and outcomes in high education. Something obvious is that assessment at this level of education should enhance the students' deep learning approach (Joughin, 2009, p. 19). Getting students to using such approach requires that the assessment tasks be well-prepared.

It should be noted that assessment has attracted and drawn the attention of many academicians and also education practitioners. Some academicians including Mardapi (2008, p. 5, 2012, p. 12) and Fook and Sidhu (2010, p. 153) account assessment as an integral or central part of teaching-learning processes. For instance, Mardapi, in that work, even goes further saying that the efforts to improve the quality of education can be reached through the enhancement of the quality of learning and the quality of its assessment system. The National Research Council [NRC] (1996, p. 5) in DiRanna et al. (2008, p. 8) also insists that assessment and learning are inseparable as they cannot be the two sides of the same coin, which means that the two are mutually inclusive.

The choice of assessment methods has balance some considerations. DiRanna et al. (2008, pp. ix–x) insist that the assessment model should balance and be susceptible (a) to effectively demonstrate how students 'represent knowledge', build knowledge in the course they are learning; (b) to display students' real performance; and (c) to be a good choice of 'an interpretation method' that allows correct inferences about students' performance. If the assessment model choice does not balance the aspects raised above, assessment may not achieve its end in education.

Fry et al. (2009, p. 198) also review how, in the beginning, researching into assessment practices in higher education was not welcome by academicians: they consider such research as either no-need-to-be-done, or as loaded of deliberate disrespect or just one way of treading down their academic space/autonomy. This can be simply considered as 'fearing the unknown' as research can lead to the

extenuating of practices that negatively affect a given educational system as Brown and Glasner (1999, p. 28) stresss it. The literature shows that research plays a lot to demonstrate to the academics that they are not geniuses not to need improvements or other new career insights.

The authentic assessment is also related to the notions of the assessor's compliance with assessment principles, formative feedback, scoring rubric, and alignment between learning activities with assessment methods, to quote but a few. It is crucial that some of these key-terms be defined in the context of this research. To begin with, assessment was defined by the University of Queensland, Australia (2007) in Joughin (2009, p. 14) as having to do with any work (which may include assignment, examination, performance or practicum) that is to be completed by a student as a requirement. Assessment is carried for different reasons, ranging from permitting the (1) grading of a student; (2) educational purposes fulfilment, like motivating students' learning, providing necessary feedback to students; and (3) as a student's official achievement record that might be availed as a proof for certification.

The afore-mentioned definition is very clear for it discloses some forms the students' tasks can take, i.e. assessment can be carried out through exams, assignments, practical tasks, and performance. It equally details that assessment has various purposes, i.e. educational and for official record about students' achievement, certifying their competence, and grading them. Educational purpose of assessment will be deepened later. More about the purpose of assessment is proposed by Irons (2008, p. 13). According to him, assessment can serve the purpose of promoting learning through providing helpful feedback, i.e. technically put, through formative assessment and formative feedback.

Feedback, as it appears in the previous line, also needs defining. It is closely related to comments on students' work in order to enhance learning and high learning achievements. According to Irons (2008, p. 13), formative feedback has to do with any piece of information, or simply a process or activity that is meant to afford or accelerate student learning and this is achieved through comments based on students' outcomes in the formative or summative assessment. The effectiveness of feedback providing depends, among other things, on whether it helps clarify what good performance is (goals, criteria, expected standards) or if it provides opportunities to close the gap between current and desired performance.

It is also important to give account of what authentic assessment is, since the whole study rotates around it. The first view is insisted by Mueller (2014) in Suarta, Hardika, Sanjaya, and Arjana (2015, p. 47) who defines authentic assessment as a form of assessment in which learners demonstrate competence, or a combination of knowledge, skills, and attitude in order to complete an essential task in a real-world situation. Based on this opinion, one can simply put that authentic assessment urges students to make use of their competence or to combine what they have already known with the existent skills just to solve a real-world problem.

Mardapi (2012, pp. 166–167) also accounts for what authentic assessment really is. Madapi stipulates that in this form of assessment, learners present or do a given assignment, the critical thinking is built in the way that students are assessed based on their ability to 'construct' or 'apply' knowledge in a real-world setting, and the evidence of what students are able to do is in live/direct, i.e. it can be observed and this turns authentic assessment to a learner-centered one. The core idea here is that authentic assessment engages students into real-world tasks that incite the use of critical thinking in constructing knowledge.

Another aspect worth underlining is that authentic assessment has got a series of methods that a teacher has to handle given the class size, the students' level of study, and ability. Teachers also smoothly use authentic assessment methods with an aim of aligning teaching-learning activities and tasks, with the assessment method chosen. Diversification of assessment techniques in authentic assessment is demonstrated in the choice offered to teachers. The latter might choose to use stu-

dents' classroom presentations, classroom discussions, individual assignments, group assessments, quizzes, examinations, students' portfolios, students' self-assessment and/or peer-assessment, projects, and performance assessment (Yusuf, 2015, pp. 292–293).

Assessment, especially in high education, is also maximally effective if it complies with a series of principle. In the Indonesian higher education context, the Ministry of Research, Technology, and Higher Education had issued principles as they can be read in the Higher Education Curriculum Book i.e. *Buku Kurikulum di Pendidikan Tinggi* (Tim Kurikulum dan Pembelajaran, 2014, p. 67). According to such a reference, any assesment should be educative, authentic, objective, accountable, and transparent.

In higher education, the literature about the tasks and course objectives alignment, and the assessment methods that enhance learning improvement and outcomes through feedback is still limited. The angle of assessment issue that is still unexploited is how the pragmatics course is assessed authentically given the role was assigned empirically to play for students who will become English language teachers. One among other reasons why only few pragmatics course assessment studies are available is given in McNamara and Roever (2006, p. 54) who comment that assessing a student's ability in pragmatics of a given language is somehow difficult. This is due to the fact that the assessor has to conciliate authentic tasks to be used and practically, given that the necessary costs required to align assessment tasks and practice are huge. However, if some researchers did not explore the angle, this does not mean it cannot be explored.

Rubrics are also great tools to be used in authentic asssessment contexts. The rubric formats used in Indonesia, indeed those mentioned in official texts about assessment, are of two types, i.e. descriptive and holistic, and lecturers may choose whichever seems comprehensible to students, efficient and effective in assessing students' knowledge, skills, and competencies. The types and formats of rubrics together with their definitions are available in the *Tim Kurikulum dan Pembelajaran*'s (2014, pp. 69–71) book, in which: (1) rubric is an assessment guide that describes the criteria used by a lecturer in assessing the result of the student's achievement level in his/her assignment/task. In addition, the rubric lists the expected performance characteristics which are manifested/demonstrated in the process and the students' work, and it also becomes a sort of reference to assess each of those performance characteristics; (2) a descriptive rubric provides descriptions of the assessment characteristics or benchmark on each given value scale; (3) holistic rubrics have only one value scale, i.e. the highest scale. The content of the description of the dimensions is the criteria of a performance to the highest scale. If the student does not meet these criteria, the lecturer comments by giving the reasons why the student cannot get the maximum score in his/her tasks.

It should be noted that the low quality of rubrics, indeed any rubric which is not clear, or simply wrongly constructed climaxes in doubts about the scoring integrity of the assessor concerned. Further, Christie et al. (2015, p. 31) investigate how assuring assessment grading tools quality affects student motivation and learning. The study displays how the Australian and USA lecturer's assessment practices of not using scoring rubrics to assess the quality of students' work tend to turn the final judgment of students' learning into a questionable one. The lecturers involved in that study tend to use common sense in assessment scoring instead of written rubrics, which could affect negatively, as the authors observed, the lecturer's integrity in grading students' work. With such conviction in mind, this study investigated the still-unexploited angle of assessment issues, that is, how pragmatics course is assessed authentically given its importance for the teacher students of English language. This research was sorely concerned with the implementation of authentic assessment in higher education. Some related aspects such as alignment, feedback, and compliance with the assessment principles are also tackled.

The problem was formulated around the idea of curiosity to know the extent to which the authentic assessment was imple-

mented in the pragmatics course taken by semester five students in the English Language and Literature Study Program. Since such an assessment has its own indicators, the problem also includes: (1) how the assessment standard is indicated in the curriculum being implemented in the pragmatics course, (2) the proof of alignment between students' tasks and the assessment methods in the pragmatics course, (3) the pragmatics course assessment methods providing more feedback to the students, (4) what the compliance with the authentic assessment principles in assessing students' tasks in the pragmatics course is like, and (5) what the authentic assessment implementation in the pragmatics course is like.

Carrying out this program evaluation was beneficial, *firstly,* to the theoretical literature by broadening it as far as the evaluation of the implementation of the authentic assessment in teaching pragmatics course to Indonesian students who are expected to be teachers of English language is concerned. Equally, this work is meant to broaden more literature regarding the use of the Discrepancy Model of Evaluation (DME) in foreign language assessment, especially in English as a Foreign Language (EFL) settings. *Secondly*, it is also beneficial to the practical aspect, because the students who are taking the pragmatics course might foster some new ideas to the pragmatics course lecturer in the perspective of adjustment as far as the course administration is concerned. Futhermore, broader space is also open to other researchers to investigate into the realms of authentic activities and assessment that might develop EFL teacher students' pragmatic competence, especially the pragma-linguistic and also socio-pragmatic competencies.

The research questions in this study were based on the problem formulated and the DEM stages, i.e. pragmatics course Program Definition, Installation, Process, and also Product (Fernandes, 1984; Fitzpatrick, Sanders, & Worthen, 2011, pp. 156–157). Those questions are: (a) to which degree did the assessment that was carried out in the pragmatics course comply with the authentic assessment standard as indicated in the curriculum? (b) what is the proof of alignment between the assessment methods used in the pragmatics course and the students' learning activities? (c) what were the most consistent feedback providing assessment methods among the ones used in the pragmatics course assessment? (d) what were the possible necessary inputs for the implementation of the authentic assessment carried out in the pragmatics course? (e) to which extent had the authentic assessment been implemented in the pragmatics course?.

Method

This research is a program evaluation that employed Provus's Discrepancy Evaluation Model. This program evaluation was carried out at a university which is located in Yogyakarta Special Region, Indonesia. The population of this study was the semester 5 pragmatics course takers. The research employed non-probability sampling method and saturated sampling technique (in which population is equal to sample) was used with n=31.

Procedure

The core is that there is a determination of: (1) the Standard (S), i.e. how the pragmatics course assessment should be conducted, based on the Ministry of Research, Technology, and Higher Education assessment principles as stated in the Higher Education Curriculum Book (*Tim Kurikulum dan Pembelajaran*, 2014, pp. 67–74), i.e. *Buku Kurikulum Pendidikan Tinggi* and the university's English Language and Literature Study Program Curriculum (2014), and then (2) taking Performance (P) measure, i.e. given the pragmatics course inputs/resources, at this stage, the pragmatics course assessment characteristics were observed, and the assessment process was scrutinised. Then, it was followed by the evaluation *per se*, i.e. the determination of discrepancies (D) by comparing Performance (P), i.e. how the program performs compared to the Standard (how it should behave).

Data, Instruments, and Data Collecting Technique

In the pragmatics course program evaluation, both quantitative and qualitative data

were collected. Three instruments were used in order to collect the data in this study, including: questionnaire, observation guide, and documentation. Through the questionnaire, the data about the assessment techniques, most feedback providing technique, compliance with assessment principles, resources, and the effectiveness of each assessment technique in uncovering the students' ability were collected. By documentation, information about the pragmatics course objectives, assessment standards, the rubrics used, and students' final learning outcomes were gathered. The observation instrument helped the authors in gathering information about the main inputs (curriculum, lecturer, and students), the assessment methods used, details about the assessment process, and teaching-learning facilities.

Data Analysis Techniques

Two types of analysis were carried out, i.e. (descriptive) quantitative analysis through Rasch Model with the Winsteps software version 3.73.0 and qualitative analysis: following Miles, Huberman, and Saldaña (2014, pp. 12–13) technique consisting of (1) data reduction or condensation, (2) data display, and (3) conclusion drawing/verification.

Evaluation Criteria

Table 1 shows the the criteria of the level of authentic assessment implementation.

Table 1. (Dis)agreement and authentic assessment level of implementation

Interval	Categories
X<-.99	Strongly Agree/Very High
-.99≤X≤0	Agree/High
0.1≤X≤1.01	Disagree/Low
X≥1.01	Strongly Disagree/Very Low

(Developed based on Sumintono and Widhiarso (2015, p. 40)
Note:
X: stands for each statement's 'Item Measure' value in logits as analysed through Winsteps version 3.73.0.

Meanwhile, Table 2 provides the information about students' scores categorization.

Table 2. Categorizing the students' scores

Score X	Categories	Criteria
X ≥ M + 1. SD	Very High	X≥ 3
M ≤ X < M + 1. SD	High	2.5≤ X <3
M - 1. SD ≤ X < M	Low	2 ≤ X< 2.5
X < M - 1. SD	Very Low	X< 2

Source: Mardapi (2008, p. 123)
Note:
M : Mean of students' final scores in the pragmatics course
X : Each single student's score out 4 (because the score scale is 4-1)
SD : Standard deviation; obtained through SD= (4-1)1/6 as the score scale is 4-1

In order to admit that a given method was used, it has to satisfy the criteria that: Mean=1 (or close to 1, that is 0.9), and STD≤0.31. Similarly, to determine whether there has been diversification of assessment methods and the students' success rate in the pragmatics course, some criteria were used:

<50% : Low
50%-65% : Average/Minimal
66%-81% : High
≥ 82% : Very High

Findings and Discussion

Before the results and discussion is presented, it should be underlined that item measure values for quantitative data are expressed in logits. For Rasch model applied in social sciences, the more the item measure value in logit gets superior to 0, the more the subjects do not agree with the statements presented to them. On the contrary, if the item measure value is equal to 0 or negative, this is an indication that the statement was agreed on by the respondents. In few words, the logit values comprised between -2 up to ≤0 are indicators that statements concerned are agreed by the respondents.

The discussion starts with quantitative data followed by qualitative data. Concerning the quantitative data, at the program Definition Stage, the resources/inputs recognized by the pragmatics course takers as primordial included: the lecturer, course objectives, classroom ability to cater for all the students, class

cleanness, sufficiency of chairs, adjustable luminosity, functional fans, and also LCD projector as their measure values in logits are respectively -0.79, -0.26, -0.57, -0.16, -0.79, -1,00, -1,00, and -1.23.

At the pragmatics course Installation Stage, the following is the comparison between the standard performance of the program and how it should behave. It is an activity aimed at finding the discrepancies. Given the pragmatics Program Process Stage/ Assessment process, the performance of the program has indicators of good performance in terms of the assessment principles of being educative, authentic, and the alignment of learning activities with the assessment used. Based on the measure values related to the positive indicators of good performance, the following measure values are more illustrative: -0.26, -0.16, -0.16, -0.16, and -0.57. It should be noted that the values represent respectively the fact that the assessment principles of being educative and authentic, the last three values are concerned with the statements about alignment.

The latter was accepted as having been observed by the lecturer of pragmatics. By doing so, she complied with the guideline which was provided in the study program curriculum, Higher education (HE) (Tim Kurikulum dan Pembelajaran, 2014), citing the Ministry of Education and Culture's Decree Number 49 of 2014, article 20, about HE in Indonesia, Sections 1 and 4 about assessment in HE.

Nevertheless, the core activity at DEM program of Installation is finding discrepancies, those which have been registered are non-compliance with the assessment principles of objectivity, accountability, and implicitly that of feedback. The item measure values associated with those three principles are superior to 0.1. The score fits to the criterion of $0.1 \leq X \leq 1.01$, so that it indicates that the respondents disagreed that there was optimization of the three principles previously mentioned. There was no use of portfolio assessment although it was recommended in the English Language and Literature Study Program and High Education Curriculum Book (*Buku Kurikulum di Perguruan Tinggi)*. As portfolio is described as a highly-recommended assessment method that allows lecturers to keep an eye on every student's knowledge process in the study program curriculum, if this lack is added to infrequency of feedback by the lecturer, the fact of not using portfolio was felt as a discrepancy.

The DEM program process stage is concerned with the results of the mostly used authentic assessment methods, the extent to which assessment methods were diversified, and the authentic assessment method, one of which is was the most feedback providing. On the list of the eleven authentic assessment methods found in the literature, six were admitted to have been used in the pragmatics course. The criteria used in determining that a given assessment method was used are that of Mean = 1, and SD ≤ 0.31. The following authentic assessment methods are satisfying: students' classroom discussion, individual assignments, quizzes, examinations, project assessment, and group assignments. The descriptive statistics (mean; SD) features are respectively: (1;0), (1;0), (0.90; 0.31), (1;0), (1;0), and (1;0). If these values are compared to the criteria pre-established, the aforementioned authentic assessment methods satisfied them thoroughly.

The second aspect looked at this point was authentic assessment method diversification. Simple calculations showed that the diversification was but average/minimal. Over the total of eleven authentic assessment methods, if six only were used, this means that the diversification was of $(6x100)/11=54.54\%$. Compared to the criteria, this percentage falls into the 50%-65% interval, which is signifying that such diversification is simply 'Average/ Minimal'.

On the top of that, the respondents' appreciation of group assignment assessments is shown in two ways: (1) they agree that it provides them with valuable feedback; (2) they recommend it to the lecturer for a better administration of pragmatics course in the future. This is indicated by its related item measure value in logits, which is -0.47. If such measure is compared to the criteria set, this illustrates that group assignments were admitted to have provided helpful feedback to the pragmatics course takers. Such finding is in

line with Bentley and Warwick (2013). Lately, students appreciate group assignment assessment as they gain learning from their friends/ peers and develop teamwork, communication, and also interpersonal skills.

Furthermore, the respondents recommend the use of group assignments, one of the techniques of authentic assessment, to the pragmatics course lecturer. This is also a case in Fook and Sidhu's (2010) study that sought to examine the implementation of authentic assessment in higher education in Malaysia, especially in the course of 'Testing, Assessment, and Evaluation 752' (TSL 752) which is taught in a Master Program at the Faculty of Education of a public university in Selangor, Malaysia. In both of these studies, authentic assessment was proven as being susceptible or appreciated to enhance learning as it won acceptance from the respondents.

Students who are successful in the pragmatics course have the scores ranging from 2.5 to 4 as it is well-described in the students' academic guide which is termed *Peraturan Akademik* (Universitas Negeri Yogyakarta, 2014, p. 15). Except for two students who were in irregular conditions, 29 out of 31 students got a score comprised between 2.66 and 4. Compared to the criteria pre-set in Table 1, students' scores fall in 'High' and 'Very High' categories.

As far as the qualitative data are concerned, the analysis led to the observation that the pragmatics course lacked clear assessment and scoring scheme, and the fact of not using portfolio although it is described as a highly-recommended assessment method that allows the lecturers to keep an eye on every student's knowledge process. The infrequency of the lecturer's feedback to students' learning and assignments was also found.

Similar findings were found in Christie et al. (2015, p. 31). Later, it is demonstrated that Australian and USA lecturer's assessment practices of not using scoring rubrics to assess the quality of the learners' work tend to turn the final judgment of students' learning into a questionable one. Simply put, if the respondents'/students' perceptions are that there was no maximization of the objectivity and accountability principles in that course, the students might have suspected the scoring integrity.

In general, the evaluation result of each stage is presented in Table 3.

Table 3. Holistic evaluation of authentic assessment implementation

No	DEM Stage/ Component	Average	Category
1	Program Definition Stage	-0.06	High
2	Program Installation Stage	-0.14	High
3	Program Process Stage	0.45	Low
4	Program Product Stage	0.02	Low
Average for the 4 DEM Stages		0.06	Low

The students' final scores in the pragmatics course are averaged and categorized as follows:

Average : 3.22
Category : Very High

Therefore, the pragmatics course definition and product (based on the students' scores aspect) are respectively in 'High' and 'Very High' categories as the average for the item measure order value for the DEM Definition stage is -0.06, while the average for the students' final scores is 3.22. The performance of the pragmatics course over the resources/inputs is also in 'High' category. Such performance is not maximal as explained by the DEM Process Stage which has the average for the item measure order value of 0.45, falling then in 'Low' category. Another aspect of the DEM product stage (concerned with the effectiveness of assessment methods used in uncovering the students' knowledge, ability, and competence) is in 'Low' category with the average for the item measure order value of 0.02.

Conclusion and Suggestions

Conclusion

A general overview of the implementation of authentic assessment is in 'Low' category. The definition and installation stages

are in 'High' category. One aspect of the pragmatics course product stage is in 'Low' category because the process itself is stained by some impediments and it is in 'Low' category. The diversification of the assessment methods is still 'Average/Minimal'. That conclusion is formulated by the following main findings. Firstly, the compliance of the pragmatics course assessment with the curriculum assessment standard is found to be in 'High' category. However, at the DEM Pragmatics Installation Stage, the discrepancies registered: (a) are little compliance with the assessment principles of feedback, objectivity, and also accountability; (b) lack the pragmatics assessment plan and scoring rubrics; (c) lack tasks and assessment methods that will push students for further research in the field of pragmatics; (d) are ineffective to support students' learning monitoring due to no use of portfolio assessment. Secondly, the proof of alignment of students learning activities and assessment methods is that: (a) the students' intended learning outcomes are in line with the study program curriculum; (b) the problem-solving skills which are engaged by the students during the learning activities resemble those required to solve assessment tasks. Thirdly, the most consistent feedback providing assessment method is group assignments. Meanwhile, the other assessment methods which are used include: (a) students' classroom discussion, (b) individual assignments, (c) quizzes, (d) examinations and also project assessment. Fourthly, the inputs which are found to be necessary for the implementation of the authentic assessment in the pragmatics course to be possible course include: (a) the lecturer, (b) the course objectives, (c) the classroom that is clean and big enough to cater for all the students, (d) enough chairs, (e) adjustable luminosity, and also (f) functional fans and LCD projector. Fifthly, the level of implementation of the pragmatics course is transcribed in the DEM Pragmatics Course Product stage that includes two aspects of the product: (a) effectiveness of the assessment methods in uncovering the students' ability, which is in 'Low' category, (b) the students' final scores in the pragmatics course, which are in 'Very High' category.

Implications

Based on the conclusions, the implications for practice are: (1) until the teachers/ lecturers choose activities that push students to use available learning resources, students will always perceive such expensive resources or services as having less importance in their learning; (2) until used up teaching/learning resources are replaced, they are seen as inexistent by students; (3) the lecturer's teaching effort and high academic competence without availing a clear assessment scheme and a scoring rubric might stain the whole scoring integrity for that teacher; (4) lecturers may use many assessment methods, and there may be alignment between students' learning activities and expected outcome assessment methods, but still assessment methods providing valuable feedback to students being very few; (5) a course where students' success rate is high as indicated by students' final scores does not implicate that the whole assessment practice has been without any spot mark.

Suggestions

Suggestions for the university administration, lecturers, and educational researchers or education practitioners are as follows. (1) The university's administration should conduct a regular check of the used-up learning resources in the classroom and replacement of those in bad conditions. (2) The pragmatics course lecturers are suggestedd to (a) apply the more student-centred teaching approach (more interactive and more chance for students to talk); (b) choose students' learning activities that push them to learn how to use resources provided by the university. (It would be unfortunate that the university presumably pays much for external journals and the Internet hotspot maintenance, but the students still say that those resources do not improve their pragmatics course learning); and (c) explain and give students opportunities to ask about either the tentative or provisional assessment scheme as well as scoring rubric. (3) Other researchers are suggested to (a) carry out other studies to evaluate the implementation of authentic assessment in the English Language and Literature Study Program particularly and all the FLA (Foreign Language Assistant) de-

partments generally, and (b) conduct other research related to the lecturers' teaching strategies/techniques, methods, and learning activities. (4) There should be a development of a model of applying Item Response Theory or any model linked to it (e.g. Rasch model) in the assessment practices in Indonesian higher education.

References

Bentley, Y., & Warwick, S. (2013). An investigation into students' perceptions of group assignments. *Journal of Pedagogic Development*, *3*(3), 11–19. Retrieved from https://journals.beds.ac.uk/ojs/index.php/jpd/article/view/199/310

Brown, S. A., & Glasner, A. (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Buckingham: Society for Research into Higher Education & Open University Press.

Christie, M. F., Grainger, P., Dahlgren, R., Call, K., Heck, D., & Simon, S. (2015). Improving the quality of assessment grading tools in Master of Education courses: A comparative case study in the Scholarship of Teaching and Learning. *Journal of the Scholarship of Teaching and Learning*, *15*(5), 22–35. https://doi.org/10.14434/josotl.v15i5.13783

DiRanna, K., Osmundson, E., Topps, J., Barakos, L., Gearhart, M., Cerwin, K., … Strang, C. (2008). *Assessment-centered teaching: A reflective practice*. Thousand Oaks, CA: Corwin Press.

Fernandes, H. J. X. (1984). *Evaluation of educational programs*. Jakarta: National Education Planning Evaluation and Curriculum Development.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines*. Boston, MA: Pearson Education.

Fook, C. Y., & Sidhu, G. K. (2010). Authentic assessment and pedagogical strategies in higher education. *Journal of Social Sciences*, *6*(2), 153–161. https://doi.org/10.3844/jssp.2010.153.161

Fry, H., Ketteridge, S., & Marshall, S. (2009). *A handbook for teaching and learning in higher education: Enhancing academic practice* (3rd ed.). New York, NY: Routledge.

Irons, A. (2008). *Enhancing learning through formative assessment and feedback*. London: Routledge.

Joughin, G. (Ed.). (2009). *Assessment, learning and judgement in higher education: A critical review*. Wollongong: Springer. https://doi.org/10.1007/978-1-4020-8905-3_2

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

Mardapi, D. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.

McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Suarta, I. M., Hardika, N. S., Sanjaya, I. G. N., & Arjana, I. W. B. (2015). Model authentic self-assessment dalam pengembangan employability skills mahasiswa pendidikan tinggi vokasi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *19*(1), 46–57. https://doi.org/10.21831/pep.v19i1.4555

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan*. Cimahi: Trim Komunikata.

Tim Kurikulum dan Pembelajaran. (2014). *Buku kurikulum pendidikan tinggi*. Jakarta: Directorate of Learning and Student Affairs, Directorate General of Higher Education, Ministry of Education and Culture.

Universitas Negeri Yogyakarta. (2014). *Buku peraturan akademik Universitas Negeri*

Yogyakarta (Revised ed.). Yogyakarta: UNY Press.

Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.

Yusuf, A. M. (2015). *Asesmen dan evaluasi pendidikan: Pilar penyedia informasi dan kegiatan pengendalian mutu pendidikan*. Jakarta: Prenada Media Group.

REiD (Research and Evaluation in Education), 3(2), 2017, 144-151
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Students' literature achievement: Predictors investigation research

Alita Arifiana Anisa
Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Karangmalang, Depok, Sleman, Yogyakarta 55281, Indonesia
Email: alita.arifiana.anisa@gmail.com

Submitted: 22 December 2017 | Revised: 06 February 2018 | Accepted: 06 February 2018

Abstract

This research is an expost facto research which aims to find out the exsistence of the mean difference between gender in terms of achievement, and investigate the variables predicting students' achievement in literature study including the direct/indirect effect. This research involved 90 students established ramdomly as the sample. The research used the quantitative data analysis to analyze the mean difference between groups and the direct/indirect effect of the predictors. The result of this research shows that: (1) girls have higher achievement in literature study compared to boys; (2) the predictors that are statistically proven as a direct significant predictor of students' final test score are gender, second dummy variable for class and mid-term test, while the rest of predictors (except the first dummy variable of class) contribute indirectly to the prediction of students' achievement in literature study; and (3) the magnitude of the predictors might be different when they are applied in different classes.

Keywords: *literature achievement, path analysis, gender, attendance*

How to cite item:

Anisa, A. (2017). Students' literature achievement: Predictors investigation research. *REiD (Research and Evaluation in Education), 3*(2), 144-151. doi:http://dx.doi.org/10.21831/reid.v3i2.17498

Introduction

As a requirement to attain a bachelor degree in an institution in Ponorogo, East Java, Indonesia, a final research project called thesis (or *skripsi* in Indonesian term) has to be conducted by students. The thesis took students' time and energy, especially in reading as well as understanding the literature needed to support their idea and findings. In order to improve the students' skill in understanding and citating relevant literature and studies, the institution conducts a compulsory course aimed at assisting the students in finding, understanding, reviewing, and citating the idea from texts. The course is called Literature Study which has to be taken by every student in the institution before they take their final research or project (thesis).

In the first semester of academic year of 2016/2017, there were ten classes consisting of approximately 350 students undertaking Literature Study. During the semester, the students have to (1) attend and actively participate in 16 face-to-face meetings, and (2) comply with all of the evaluation requirements, including oject presentation and mid-term test. The issue is that although the subject is considered to be very important for the success of their final research/project, the students seemed to think that the subject was not as important as their main subjects (the subjects directly connected to their major), so their achievement in the course was not satisfactory. The unsatisfying final score leads to the stakeholders' anxiety related to the quality of the students' final research/project (thesis). The poor quality of their thesis is one of the indi-

cations of the poor academic ability in integrating all of the knowledge that the students have earned in their four-year study.

Phye (1995) states that learning and achievement are surely related, but they are different in significant ways. People start to learn anything consciously or even unconsciously to achieve their desired objectives. For example, when you saw someone playing a doll-fishing machine and found that he never missed the doll, you were excited to know how he did such a good thing and started to observe every single thing he did in order to catch the doll. The observation you did to gain any information in doing doll fishing is a learning process and to be a good doll-fisher is your objective. After gaining information, you start to test whether the information works by challenging yourself to do doll fishing by yourself, and the result of your doll-fishing test represents your achievement. It could be good (you were successful in catching the doll by using the information you have got from the learning process, i.e, observation), but it could also be not too good (you missed the doll). The achievement (either to be good or not too good) shows whether the learning process you did earlier meets your objectives.

Pritchard (2009), in *Ways of Learning*, mentions some definitions of learning, including: (1) a change in behaviour as a result of experience or practice, (2) the acquisition of knowledge, (3) knowledge gained through study, (4) gaining knowledge of, or skill in, something through study, teaching, instructions, or experience, (5) the process of gaining knowledge, (6) a process by which one's behaviour is changed, shaped or controlled, and (7) the individual process of constructing understanding based on the experience from wide range of sources. Thus, learning is a process in gaining knowledge through experiences and proven by behavioural changes. The experience means any kinds of experience, including educational experience through teaching and learning processes at school.

Meanwhile, achievement is what you gain from learning processes. The definition of academic achievement by the *Dictionary of Education* in Phye (1995) is an accomplishment or proficiency of knowledge or skill. The achievement shows the increase of the learners' knowledge after experiencing learning. One of the ways to see the learners' achievement is by seeing their changes. In order to know their achievement as a result of learning process, teachers need to conduct assessment. American Educational Research Association (AERA, 1999) in Reynolds, Livingston, and Willson (2009) mentions that in general, assessment is any systematic procedure to collect the information to make inferences about the characteristics of people. In educational issue, assessment can be defined as a procedure to gain any information about students' learning or value judgement concerning learning process through observations, ratings of performance, project or tests (Miller, Linn, & Gronlund, 2009).

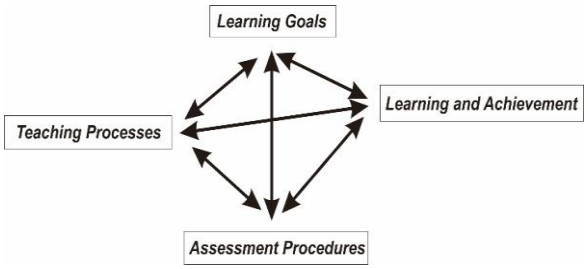


Figure 1. The relationship among learning, achievement, and assessment (Cumming & Maxwell, 1999)

There are some procedural questions which need to be answered in conducting an assessment: First, what are the learning objectives/goals which need to be achieved? Learning leads to the changes of knowledge. Thus, the first step that the teacher needs to do to assess students' learning achievement is deciding the specific learning objectives or identifying what the teacher wants the students to master after the learning process. Although Mueller in Berg (2006) mentions that it is not easy for a teacher to write good learning objectives, it is essential to acquire a clarity of purpose, because with clear purposes in mind, assessment can be well designed to match the purposes (Phye, 1995).

Second, what kind of assessment approaches matches the learning objectives? Identification to determine the type of goals needs to be held. Is it cognitive changes

which come as the result of content acquisition? Is it motor changes in performing specific task? Or is it behavioural changes? The identification process is important to help teachers to choose the best assessment approaches and tools to meet their teaching objectives. The assessment approaches to each learning objective are as follows. (1) Cognitive objectives include the building of knowledge base (Thorndike & Thorndike-Christ, 2010). In order to meet the cognitive changes as a result of content acquisition type of learning objectives, various types of test can be conducted, such as multiple-choice item, matching, true or false, essay, short answer or filling in the blank tests (Phye, 1995). (2) Performance objectives cover the motor changes and how the learners perform their knowledge in the form of action/skill in doing a specific task. In order to match the performance objectives, a task that requires students to demonstrate a specific action or project is appropriate, such as playing musical instrument, using software to analyse data, constructing housing model, and making financial report. (3) Affective/behaviour objectives involve the development of attitudes, values, interest and personal or social attributes that teachers can assess through observation (and try to infer what lies behind the behaviour), peers' or teachers' rates, and also students' self-reports (Thorndike & Thorndike-Christ, 2010).

Third, after the set of operations that requires students to perform their cognitive, performance, and attitude changes has been accomplished, it is important to set the rule to value students' responses. The rule which is called scale is crucial to decide the most suitable number that is able to represent how much the objective is existing (Thorndike & Thorndike-Christ, 2010). There are four kinds of scale which are used in measurement theories, namely: nominal scale (the number on the scale does not refer to the amount of anything), ordinal scale (the number on the scale tells the order of specific condition without knowing how much something is less or more than something else), interval scale (each number on the scale has equal difference, zero in this scale is not an absolute zero, and ratio scale (the scale that has an absolute zero).

The educational objectives are measured by various types of instruments (the assessment tools) that are able to cover cognitive, performance, and affective objectives. The instrument is assumed to have equal amount of traits in every item. The equal differences in score indicate the equal differences in traits, and, thus, they fulfil the requirement to use interval scale. In the interval scale, the absolute zero does not exist; it is suitable to the educational issues where the students are not assumed as an empty vehicle (the base knowledge existence assumed). After setting the specific rule to value students' responses, the next step to do is scoring. The common mechanism to do a scoring is by calculating the relative achievement objective. Relative mastery involves estimating the percentage of the domain that the students have mastered. For example, when students answer eight out of ten questions correctly, it indicates that the students have mastered 80% of the domain (Thorndike & Thorndike-Christ, 2010).

The study related to the variables which can predict the final test score is beneficial to provide the stakeholder (lecturers and academic authorities) of the institution an evidence to formulate the decision to perform a better package of treatments to assist the students in preparing themselves to face the final research through Literature Study course. By the study, the lecturers are able to decide what to be focused in order to optimize students' literature achievement. The main purpose of the study is to identify the variables that are able to predict students' achievement in Literature Study.

The study is significant since there has never been a study related to the final test score in Literature Study in Ponorogo, although the literature study is considered to be beneficial to students in conducting their final research/project (as one of the requirements to graduate from the bachelor degree). This study is able to provide the lecturers and academic authorities an empiric evidence to give an appropriate treatment to help the students to reach their optimum achievement.

Method

Population and Sample

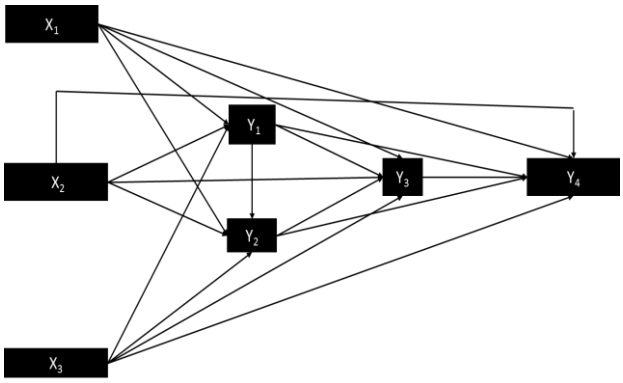
The population of the study was all of the students who are majoring in Islamic religion education. They took literature study course in the academic year of 2016/2017. The sample of the study was 90 students from three classes of X, Y, and Z who are chosen randomly.

Data Variables

The data which were employed in this research included: (1) gender, (2) classes, (3) attendance, (4) project presentation score, (5) mid-term test score, and (6) final-test score or achievement. Table 1 shows the description of the data.

Research Procedures

This research is an expost facto research which studies about the variables that occured in the past. The research employed the quantitative research approach in order to investigate two independent variables and also four dependent variables. The research covered: (1) a descriptive analysis for analyzing the frequency of the categorical data (gender and classes) and the central tendency, variance, standard deviation, skewness and kurtosis of the continuous data (attendance, project presentation, midterm test, and final test); (2) mean difference to analyze the mean difference between genders in achievement; (3) a path analysis to analyze the direct and indirect effect in predicting the independent variable, its equation and the final model; and also (4) a path analysis for each class. Meanwhile, the hypothesis of the research model analyzed is presented in Figure 2.



X1: Gender; X2: Classdummy1; X3: Classdummy2
Y1: Attendance; Y2: Project Presentation; Y3: Mid-term Test
Y4: Final Achievement

Figure 2. Hypothesis research model

Findings and Discussion

Findings

The sample consists of 90 participants, more than half (64.44%, n = 58) of the participants are girls, and 35.56% of the samples are boys (n = 32). The samples are the students who are studying literature study in three different classes, in which 27.8% (n=25) are class X students, 37.8% (n=34) are class Y students, and the rest 34.4% (n=31) are class Z students.

Table 1. Data descriptions

Data	Description	Data Type	Data Source
Gender	1 = Male 0 = Female	Categorical	Student identity document
Classes	1 = X 2 = Y 3 = Z	Categorical	Academic document
Attendance	The students' attendance in 16 meetings.	Continues	Teacher-attendance report
Group project presentation	The average score of group and individual performance The score in 1 to 100.	Continues	Student performance report
Mid-term test score	The score is 1-100.	Continues	Mid-term test results
Final test score	The score is 1-100.	Continues	Final test result

Figure 3 shows the distribution of the variables which are studied in this research. All of the variables are considered to be normally distributed, with the skewness of less than ±2.0, and kurtosis of less than ± 7.0. Significantly, the data indicate that the attendance variable is ranging from 57 to 100 (M=93.91, s= 7.523), while the project variable ranges from 52 to 89 with the average score (M) of 79.74 and standard deviation (s) of 5.867. Furthermore, the data also show that the average score of the students' mid-test score is 71.10 (which is ranging from 50 to 98, s= 11.138), and the final test score is ranging from 50 to 100 (in which M=78.17, s=12.002).

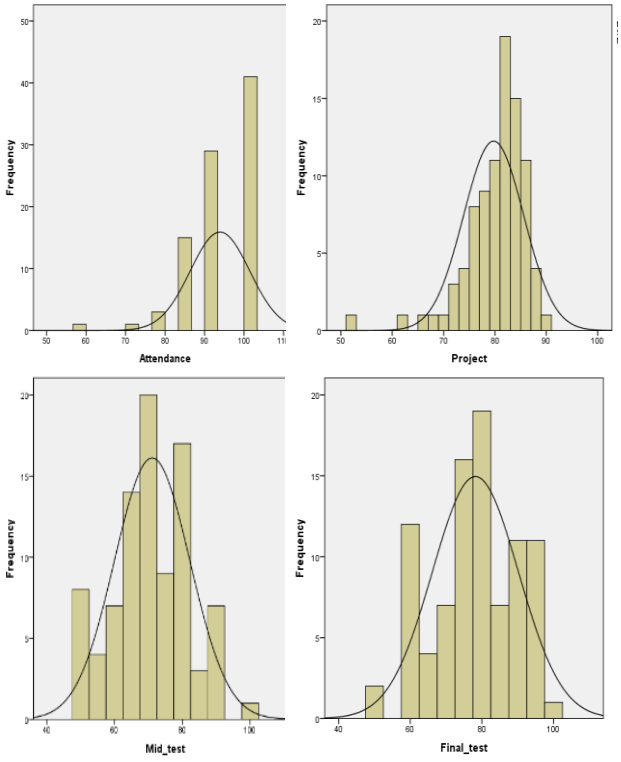


Figure 3. The distribution of the attendance, project, mid-term test and final test

Analysis of Mean Difference

Gender to Achievement. The independent sample t-test was conducted in order to reveal whether there is a significant mean difference between boys and girls in the Literature Study course achievement. The results indicate that girls have a greater mean score in the Literature Study achievement. Table 2 presents the result of the mean difference analysis.

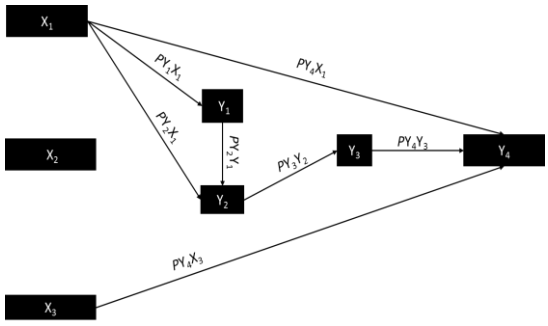
Table 2. Mean difference analysis result

Variable	Gender	N	Mean	S	T	Sig
Final	Girls	58	73.03	11.12	-3.74	.00
	Boys	32	72.22	10.98		
	Girls	58	81.45	11.34		

N= 90, p <.05

Path Analysis

A regression analysis was conducted in order to investigate the predictors of attendance, project presentation, mid-term test, and also final test (dependent variable). There were two independent variables: gender and class. Both of the independent variables were categorical data; gender consisted of two categories (boys and girls), while class consists of three categories (X, Y, and Z). The predictor that consisted of three categories could not be simply categorized into 0 and 1, so that a dummy variable needed to be created. A dummy variable is a way to represent groups of people or condition using only zero and one, and the number of dummy variables is one less than the number of the groups recoded (Field, 2013). The final model is shown in Figure 4.



P= Path Coefficient

X1: Gender; X2: Classdummy1; X3: Classdummy2
Y1: Attendance; Y2: Project Presentation; Y3: Mid-term Test; Y4: Final Test

Figure 4. Research's final model

Pedhazur (1997) explains that in simple regression, β is equal to correlation coefficient (*r*). He also mentions that the path coefficient from variable 1 to variable 2 is equal to β_{21}, which can be estimated from the data by calculating r_{12}. Thus, in this research, the coefficient for each path is clearly presented in Table 3.

Table 3. Path coefficients

Path Coefficient (*P*)	Unstandardized		Standardized (β)
	A	B	
PY_1X_1	95.414	-4.226 X_1	-.270 X_1
PY_2X_1	47.513	-2.430 X_1	-.199 X_1
PY_2Y_1		+ .352 Y_1	.452 Y_1
PY_3Y_2	15.146	+ .702 Y_2	.370 Y_2
PY_4X_1	54.627	- 7.685 X_1	-.308 X_1
PY_4X_3		-6.499 X_3	-.259 X_3
PY_4Y_3		.401 Y_3	.372 Y_3

Gender and Classes on Attendance. The result of the multiple regressions using the stepwise method to investigate the predictors of Attendance shows that gender is the only independent variable that is statistically significant in predicting the attendance. There are 7.3% attendance variances explained by gender. The equation used to predict students' attendance is as follows:

$$\mathbf{Y_1 = PY_1X_1}$$

Gender, Classes and Attendance on Project Presentation. Investigation was conducted using multiple regressions to find out the variables which predict project presentation. The result shows that 29% variances of project presentation are accounted by gender, classes, and attendance. The analysis found that gender and attendance are statistically significant to predict the students' project presentation score (<.05), while classes are not significant (>.05).

Unlike the previous multiple regression equation, the equation used to predict the students' project presentation considered both direct effect (gender on project presentation) and indirect effect (gender on project presentation through attendance) using the path analysis. In the path analysis, the sum of the direct effect and indirect effect is called the total effect, or effect coefficient (Pedhazur, 1997). The equation to predict the students' project presentation is as follows:

$$\mathbf{Y_2 = PY_2X_1 + (PY_1X_1 * PY_2Y_1)}$$

Gender, Classes, Attendance, and Project Presentation on Mid-term Test. Another multiple regression analysis was conducted in order to investigate the predictors of students' mid-term test scores. The analysis result indicated that gender, classes, attendance, and project presentation are able to explain 13.7% variances of students' mid-term test score, but only project presentation is statistically significant in predicting mid-term test (<.05). The equation which was used to predict the students' mid-term test score has no direct effect, so the equation is constructed only by considering the indirect effects of: (1) gender on mid-term test through project presentation, and (2) gender on mid-term test through attendance and project presentation. The equation is as follows:

$$\mathbf{Y_3 = (PY_1X_1 * PY_3Y_2) + (PY_1X_1 * PY_2Y_1 * PY_3Y_2)}$$

Gender, Classes, Attendance, Project Presentation, Mid-term Test on Final Test. The last multiple regressions conducted was to find out which independent variables (gender, classes, attendance, project presentation, and mid-term test) are able to predict students' final test score. The result shows that there are 33.4% variances of students' final test scores which are accounted by gender, classes, attendance, project presentation, and mid-term test result, but only mid-term test, gender and also classdummy2 (second dummy coding variable for classes) that are statistically proven as a significant predictors of students' final test scores. The equation was constructed by considering the direct effects (gender on final test and classdummy2 on final-test) and indirect effects (gender on final test through project presentation and mid-term test, and also gender on final test through attendance, project presentation and mid-term test). The equation which is used to predict students' final test score is as follows:

$$\mathbf{Y_4 = PY_4X_1 + PY_4X_3 + ((PY_2X_1 * PY_3Y_2 * PY_4Y_3) + (PY_1X_1 * PY_2Y_1 * PY_3Y_2 * PY_4Y_3))}$$

Predicting Final Test Score in Different Classes. Every class has its own characteristics influenced by the students' environmental and academic background. The idea of the differ-

ent characteristics of class leads to a further analysis to compare the contribution of each predictor in different classes. Table 4 shows the result of the multiple regressions using the enter method in each class.

Table 4. Comparing the effects of gender, attendance, project, mid-term test on final test in different classes

Predictors	Final Test		
	X	Y	Z
Gender	-.502*	-.171	-.275
Attendance	.067*	.228	-.388
Project Presentation	-.122	.102	.386
Mid-term test	.399*	.348	.317
R_2	.495	.434	.337
E	.505	.566	.663
N	25	34	31

P<.05

The result shows that in Class 1 (X), the biggest predictor is gender and the lowest one is attendance. In class Y, the biggest predictor of the students' final test is mid-term test score, and the lowest predictor is project presentation. In class Z, the biggest predictor of students' final test is attendance, while the lowest one is gender. The result proves that the predictors might predict the independent variable in different magnitude based on the class characteristics. Due to the different magnitude of the predictors, it is recommended that lecturers treat the classes differently.

Discussion

The research findings show that girls have a significantly greater mean score than boys in terms of Literature Study achievement. In line with the mean difference analysis result, the final model of this study also shows the contribution of gender to students' final achievement. Two previous researchers, Downing (1977) and Droege (1967), mention that girls have greater facility in early reading skill. The evidence of girls' reading ability is also revealed by Finucci, Gottfredson, and Childs (1985). Based on the reserach, women become better oral readers, buy more books, and read more pleasure than men.

Beside the direct contribution to students' achievement, gender is also found to be the predictor of students' attendance rate and project persentation performance. Based on the lecturer reports, girls tend to attend the class more often than boys, and girls are more serious in doing their project persentation homework. In line with this, the research of Duckworth and Seligman (2005, p. 939) also explains that girls are more self-disciplined than boys in final grades, school attendance, and hours-spending homeworks.

Unlike gender, the first dummy variable (X=0, Y=1, and Z=0) has no significant effect on any independent variables. The second dummy variable (X=0, Y=0, and Z=1) has a direct negative significant effect on students' achievement. It means that class Z has significantly lower achievement compared to the other classess (X and Y). Based on the predictor analysis conducted for each class, the low achievement of class Z was due to the low score of students' attendance.

Students' attendance, which is influenced by gender, is statistically proven as the predictor of the students' project presentation performance. Attendance is also found to be the indirect predictor of final achievement through the project presentation and mid-term test score. The previous research also found a similiar phenomenon. The research of Deane and Murphy (2013) found that the students' attendance is positively correlated with overall examination score. In addition, Louis, Bastian, McKimmie, and Lee (2016) also found the positive correlation between attendance and objective performance.

Conclusion and Suggestions

Conclusion

Based on the findings, conclusions can be drawn as follows: (1) girls have higher achievement in Literarture Study compared to boys; (2) the predictors which are statistically proven as direct significant predictors of students' final test score are gender, second dummy variable for class, and mid-term test, while the rest of the predictors (except for the first dummy variable of class) contribute indirectly to predicting students' achievement in Literature Study; (3) the magnitude of the predictors might be different in different classes.

Limitation of the Research

The limitation of the research is that: (1) the research was conducted in an Islamic Institution, in which the references studied in the Literature Study course are those related to the Islamic religion education, prophetic character, and intellectual character; (2) every institution has their own policy related to what kind of academic activities that the students have to pass through during Literature Study course in one semester.

Suggestions

To achieve optimum final test score in Literature Study, the lecturers are suggested to: (1) consider gender, class, attendance, project presentation score, and mid-term test score to improve students' final test score; (2) be concerned with the mid-term test results, because only mid-term test has a direct effect on the final test score; and (3) treat each class differently based on their own characteristics.

References

Berg, S. L. (2006). Two sides of the same coin: Authentic assessment. *Community College Enterprise*, *12*(2), 7–21. Retrieved from https://www.questia.com/read/1P3-1167542141/two-sides-of-the-same-coin-authentic-assessment

Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policy & Practice*, *6*(2), 177–194. https://doi.org/10.1080/09695949992865

Deane, R. P., & Murphy, D. J. (2013). Student attendance and academic performance in undergraduate obstetrics/gynecology clinical rotations. *JAMA*, *310*(21), 2282–2288. https://doi.org/10.1001/jama.2013.282228

Downing, J. (1977). How society creates reading disability. *The Elementary School Journal*, *77*(4), 274–279. https://doi.org/10.1086/461058

Droege, R. C. (1967). Sex differences in aptitude maturation during high school. *Journal of Counseling Psychology*, *14*(5), 407–411.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939–944. https://doi.org/10.1111/j.1467-9280.2005.01641.x

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Los Angeles, CA: Sage Publication.

Finucci, J. M., Gottfredson, L. S., & Childs, B. (1985). A follow-up study of dyslexic boys. *Annals of Dyslexia*, *35*(1), 117–136. https://doi.org/10.1007/BF02659183

Louis, W. R., Bastian, B., McKimmie, B., & Lee, A. J. (2016). Teaching psychology in Australia: Does class attendance matter for performance? *Australian Journal of Psychology*, *68*(1), 47–51. https://doi.org/10.1111/ajpy.12088

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching* (10th ed.). Upper Saddle River, NJ: Pearson.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Orlando, FL: Harcourt Brace College.

Phye, G. D. (Ed.). (1995). *Handbook of classroom assessment: Learning, achievement, and adjustment*. Ames, IA: Academic Press.

Pritchard, A. (2009). *Ways of learning: Learning theories and learning styles in the classroom* (2nd ed.). Oxford: Routledge.

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

Thorndike, R. M., & Thorndike-Christ, T. M. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Pearson Education.

REiD (Research and Evaluation in Education), 3(2), 2017, 152-162
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Characteristics and equation of accounting vocational theory trial test items for vocational high schools by subject-matter teachers' forum

Dian Normalitasari Purnama
Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
Email: diannsp@gmail.com

Submitted: 23 January 2018 | Revised: 12 February 2018 | Accepted: 12 February 2018

Abstract

This study is aimed at: (1) understanding the characteristics of Accounting Vocational Theory trial test items using the Item Response Theory and (2) determining the horizontal equation of Accounting Vocational Theory trial exam instruments. This was explorative-descriptive research, observing the subject of the eleventh-grade students. The research objects were test instruments and responses of students from six schools selected through the stratified random sampling technique. The data analysis employed review sheets and BILOG program for the Item Response Theory 2PL. The findings were as follows. (1) The test item review of test packages A and B found 37 good quality items, the Item Response Theory using 2PL showed that Package A Test generated 27 good questions, Package B Test contained 24 good questions. (2) The question equating using the Mean/Sigma method resulted in the equation of b_2^*= 1.168bx + 0.270, with the Mean/Mean method resulting in the equation of b_2^* = 0.997bx - 0.250, the Mean/Mean method at 0.250, while Mean/Sigma method at 0.320.

Keywords: *accounting questions, vocational high school, horizontal equating, Item Response Theory*

How to cite item:

Purnama, D. (2018). 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. doi:http://dx.doi.org/10.21831/reid.v3i2.18121

Introduction

Nitko and Brookhart (2011, p. 3) define assessment as a broad term referring to a process for obtaining information used for making decisions about students; curricula, programs, and schools; and educational policy. Assessment and evaluation of learning outcomes are among the efforts made to monitor the students' competency following the learning process. In accordance with Article 57 Paragraph (1) of Law No. 20 of 2003 on National Education System, evaluation is performed in the national education quality control framework to show education provider's accountability to interested parties such as students and educational institutions and programs. The evaluation, for instance, is implemented by the government through National Examination (*Ujian Nasional* or UN).

National Examination is held annually and simultaneously across Indonesia. Regulation of the Minister of Education No. 20 of 2007 on the educational assessment standard explains that National Examination is an activity which measures students' competency in certain science and technology subjects to appraise their achievements in National Education Standards. The outcomes of the National Exam are further used by the government to establish policies pertaining to education. Article 68 of Government Regulation No. 19 of

2005 on National Education Standard mentions that the outcomes of the National Exam are used as a consideration in mapping the quality of educational program and/or unit. The mapping has the purpose to understand the quality of education in each region.

Before National Examination is held, the Provincial and Regency Education Offices hold trial exams (nationally known as 'try-outs')- as a preparation for students in facing the exam. In an interview between the researcher and an accounting teacher at a vocational high school, the teacher said that he chaired the Accounting Subject-Matter Teachers' Forum (*Musyawarah Guru Mata Pelajaran* or MGMP) of Sleman Regency. The interview revealed that the test used in the accounting trial exam for vocational high schools held by the Education Office of Sleman Regency, particularly for the Productive Accounting subject, was prepared by the Accounting Subject-Matter Teachers' Forum. The questions were given in two packages (A and B), with the same exam content outline and materials to avoid cheating during the trial exams.

Both packages for the Accounting Vocational Theory trial exam for vocational high schools in Sleman Regency can be used as a collection of questions with good characteristics. A good test instrument is composed of good items (Retnawati, 2014, p. 62). Therefore, an analysis of test items contained in a test instrument is necessary to help finding out the quality of the instrument. Mardapi (2012, p. 128) suggests that an item analysis can observe the difficulty level, discrimination index, and distractor's effectiveness of test items. The analysis also helps in observing the validity and reliability of a test.

In addition to test item characteristics, the parallelism of both trial test packages is unproven. This means that the difficulty level and discrimination index of both test packages may or may not be the same. This can cause a student's scores to be higher than his ability, and thanks to the easier test package he received. This situation may result in the inaccurate measurement in students' competency achievement. For this reason, although both packages for the Accounting Vocational Theory trial exam prepared by the Accounting Subject-Matter Teachers' Forum are provided with the same exam content outline and materials, the equation between package A and B still becomes a subject of attention.

When the parallelism of the two test packages is proven, an equation process is the next step to be taken. Kolen and Brennan (2014, p. 2) define equation or equating as a statistical process in order to adjust the scores of a test so that they can be used interchangeably. Sukirno (2007) explains that equating can compare the scores earned by students albeit using different test packages. In that way, test participants will not be disadvantaged by easier or harder test packages they receive. There are two approaches that can be used for test equating: Classical Test Theory (CTT) and Item Response Theory (IRT). In CTT, the test to be equated must have the same reliability index. The Item Response Theory, which utilizes the mathematical model, determines that the probability of test participants in giving the right answer to a question depends on the ability they possess and also the characteristics of the question (Hambleton, Swaminathan, & Rogers, 1991, p. 9).

Test equating using IRT is more representative than that using CTT, since IRT has invariance characteristics in its parameter. The ability parameter is invariance with the test parameter and vice versa (Aminah, 2012). The same measurement scale in the scores obtained by students during a trial exam will make education quality monitoring easier. The test outcomes will show the students' competence mastery in facing the National Exam, while serving as a consideration for making decisions for improving the quality of graduates.

Hambleton and Swaminathan (1985, p. 197) explain that horizontal equating is performed between two different versions of a test, and vertical equating is performed on tests across the difficulty levels. Horizontal equating can also be defined as determining the equal score for differences (Crocker & Algina, 2008, p. 456). Horizontal equating is proper when it is used for the security of a test, so that several forms of tests are needed. These forms are not the same, but it is expected that they are similar in their content and difficulty. When the difficulty, reliability, and content of

tests are so different from one form to another, few methods of equating can properly work (Cook & Eignor, 1991). Dorans, Moses, and Eignor (2010) mention that in an equivalent group design, two tests are administrated to two equivalent groups chosen randomly from the same population (they are assumed to have equivalent ability). Moghadamzadeh, Salehi, and Khodaie (2011) also explain that the equivalent group design might reduce the effect of exercise and boredom, but it might also cause a bias since they might not have equivalent distribution of ability. To reduce the possibility of bias, the use of a big sample is suggested. In addition, Liao and Livingston (2012) present three approaches that could be considered as alternatives to a common-item equating design. In their paper, the randomly equivalent form approach assembles the test forms of equal difficulty by stratified random sampling of items from the item pool. Previous study which was conducted by Miyatun and Mardapi (2000) also introduces the non-anchor item equating using the equivalent group design.

The above description illustrates the significance of equating both test packages of Accounting Vocational Theory trial exam for vocational high schools prepared by the Accounting Subject-Matter Teachers' Forum of Sleman Regency. The question analysis and test instrument equating will realize objective information and show the actual competency of students in preparing for the National Examination.

Method

This descriptive-quantitative research tries to equate the test instruments of Accounting Vocational Theory trial exam for vocational high schools that were prepared by the Accounting Subject-Matter Teachers' Forum of Sleman Regency in the academic year of 2015/2016 in two packages, A and B. The research was conducted at vocational high schools in Sleman Regency, Yogyakarta Special Region. The subjects of this research are grade XII students of vocational high schools in Sleman Regency who took the Accounting Vocational Theory trial exam in the academic year of 2015/2016. The research objects were test instruments and also 650 students' Package B participants in the form of answer sheets from six vocational high schools selected through the stratified random sampling technique based on the National Exam rank for Accounting Vocational Theory subject in the academic year of 2014/2015.

Kolen and Brennan (2014, p. 13) state that there are two ways to do an equivalent group design: (1) by giving single test to measure students' ability, and (2) by doing a structure test administration, for example, X test for the first student, Y test for the second student, X test for the third student and so on. In reference to the theory, the accounting competency try out test is considered to be suitable with the equivalent group design since the students with odd number of students identity were working with Package A test, while those with even number working with Package B test.

The data were collected through documentation. They were reviewed by experts to see the characteristics of the test items qualitatively. The review of the test items was made to material, construction, and language to see their qualitative characteristics. The trial exam answer sheets or responses were used for the quantitative analysis. The test instruments were analyzed using the Item Response Theory with the assistance from the BILOG-MG program to generate three-phase output. In the first phase, it revealed the number of test participants correctly answering test items, ratio of correct answer probability divided by wrong answer probability, and biserial coefficient. The second phase obtained the data on item parameter according to the Item Response Theory model used. The 1-PL model covers the data on the difficulty level, the 2-PL model covers information on the difficulty leve, and discrimination index, and the 3-PL model covers the difficulty level, discrimination index, and guessing factor. In estimating the parameter, the logistic model with the highest number of fit items was used. Fit items are items with calculated Chi-Square value smaller than table Chi-Square value or *p-value* above 5%. The goodness of fit test aims at knowing whether or not the items used are in accordance with the model applied.

The level of difficulties is an item category, easy or uneasy item to students. It can be understood by calculating the number of students who answer correctly. It is considered good when the scores range from -2 to 2, the discrimination index is considered good when the scores range from 0 to 2, and guessing factor is considered good when the score is lower than 0.2 (1/total answer alternatives).

The testing of the equation of the two test packages is aimed at observing whether or not Packages A and B tests were parallel. In the presence of any evidence of non-parallelism, both packages need to be equated. Allen and Yen (1979, p. 59) suggest that two test instruments are considered parallel when both have the same mean and variance. The parallelism testing of the test instruments was carried out using the SPSS Program.

Equating was carried out based on the result of parameter estimation from BILOG which generates information on the equated test instrument conversion constant. Equating was held using equivalent group design since, as shown by the data, the students' responses were sourced from two different test instruments and answered by two different student groups with equivalent ability. There were no anchor items in both test instruments.

Findings and Discussion

Findings

Validity and Reliability of Questions

This research involved five raters to estimate the validity using Aiken formula. The validity of the test items in both Packages A and B according to Aiken formula is relatively good. Package A contains 26 questions with good validity index (minimum 0.87) and also 14 questions with poor content validity. Package B contains 27 questions with good content validity index. There are 13 questions with very poor content validity index.

Characteristics of Accounting Vocational Theory Trial Test items based on Question Item Review Criteria

Table 1 shows the characteristics of trial test items based on the outcome of expert review. In the material aspect, test Packages A and B have 37 good questions and three poor questions. This is due to the reason that the prepared questions are not in accordance with the exam content outline. In the material and language aspects, 40 items in both Packages A and B are in a good category.

Characteristics of Accounting Vocational Theory Trial Test items based on Item Response Theory

The quantitative analysis using Item Response Theory requires an assumption test as a prerequisite. A unidimensional assumption test was carried out to observe whether or not the Accounting Vocational Theory trial exam instruments measure one's ability (trait). The unidimensional test was performed with the factor analysis using SPSS 20. As presented in Figure 1, the result of the factor analysis shows that 40 test items form 11 factors that explain 55.063% of the total variance. The result also shows that the first factor is dominating, with Eigen value of 9.439 which is five times bigger than the second factor. Therefore, it is safe to say that Package A of the Accounting Vocational Theory trial exam instrument is unidimensional.

Table 1.Outcome of trial test items review

Aspect	Package	Question Criteria					
		Good		Poor		Very Poor	
		Qty	%	Qty	%	Qty	%
Material	A	37	92.5	3	7.5	-	-
	B	37	92.5	3	7.5	-	-
Construction	A	40	100	-	-	-	-
	B	40	100	-	-	-	-
Language	A	40	100	-	-	-	-
	B	40	100	-	-	-	-

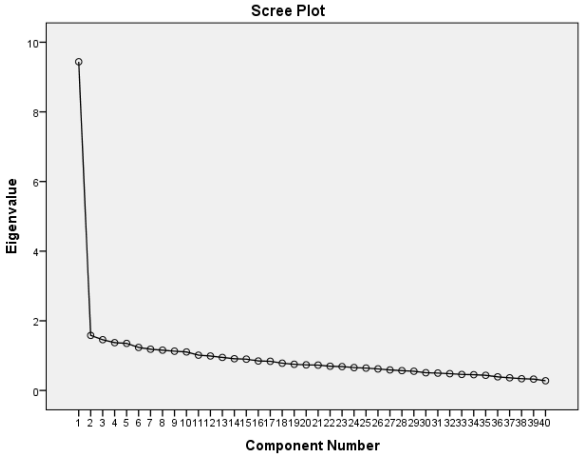


Figure 1. Scree plot of Package A

As presented in Figure 2, in Package B Test, 40 test items form 13 factors which explain 56.740% of the total variance. The result also shows that the first factor is dominating, with the Eigen value of 7.595 which is four times larger than the second factor. Therefore, it can be assumed that Package B of the Accounting Vocational Theory trial exam instrument is unidimensional.

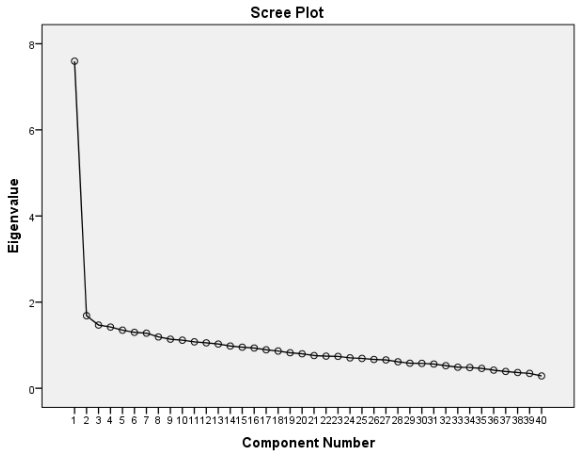


Figure 2. Scree plot of Package B

Local independence assumption test for Package A is proven with variance-covariance matrix and students' ability in doing Package A test, where the students were divided into 15 groups. The classification was carried out by listing the students' rank from the highest to lowest ability. The classification was held using the 2-parameter ability estimation model. The result shows that the elemental value is outside the diagonal approaches, meaning that the test instruments have passed the local independence assumption test.

The parameter invariance assumption test came in two types. The first was question item parameter invariance test which is aimed to observe whether or not the test questions changed when answered by different student groups. The second was parameter invariance test on participants' abilities to see whether or not the estimated students' abilities changed when the test items were changed. The test was performed using scree plots as presented in Figure 3, 4, and 5.

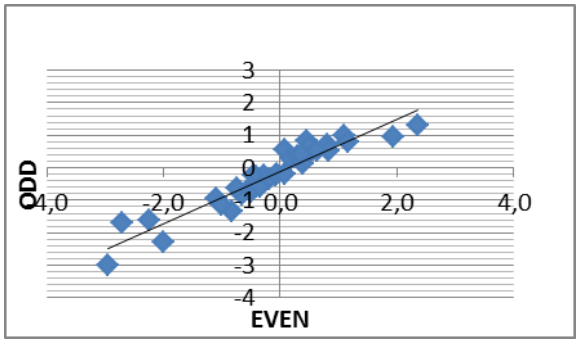


Figure 3. Scree plot of parameter invariance for the difficulty level in Package A test

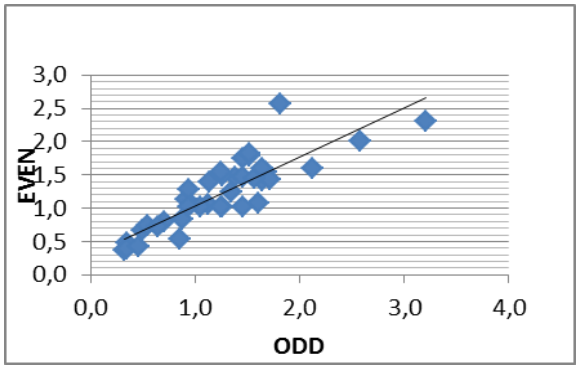


Figure 4. Scree plot of parameter invariance for discrimination index in Package A test

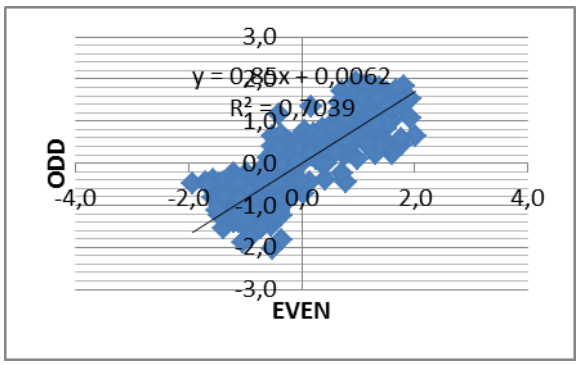


Figure 5. Scree plot of parameter invariance for participants' abilities in Package A test

Figure 3, 4, and 5 show that in general, all of the plots are relatively close to the diagonal line, which can be read that the parameter invariance in Package A Test is met.

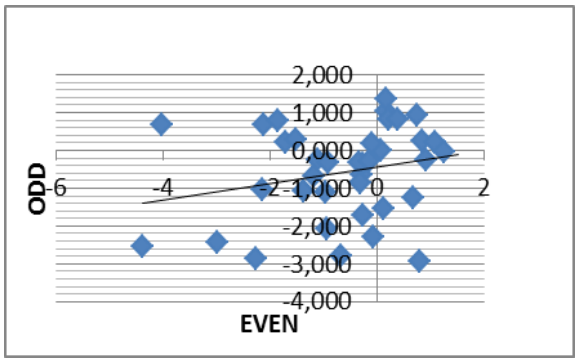


Figure 6. Scree plot of parameter invariance for the difficulty level in Package B test

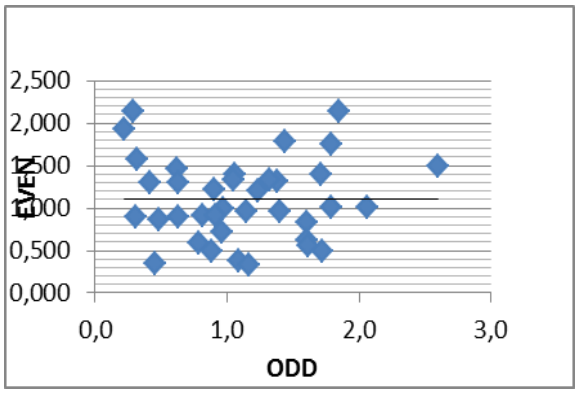


Figure 7.Scree plot of parameter invariance for discrimination index in Package B test

Meanwhile, figure 6 and 7 show that in general, all plots are scattered, away from the diagonal line. Scattered plots away from diagonal line show that the invariance parameter of the difficulty level and the discrimination index of Package B test are not met.

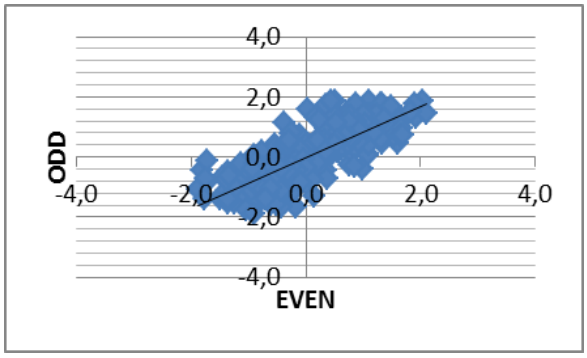


Figure 8. Scree plot of parameter invariance for participants' abilities in Package B test

Figure 8 shows that, in general, all plots are relatively close to the diagonal line. Therefore, it can be inferred that the assumption for invariance parameter for students' abilities in Package B Test is met.

The Result of Model Fitness. In order to determine the model that is fit to the items, data analysis under the three parameter logistics was conducted (1PL, 2PL, and 3PL). The fit-model analysis was assisted by BILOG software version 3.0. The fit-items were the items with Chi-Square value bigger than 5%. The fit-model analysis was beneficial to the determination of the model fitness test to this modern approach by using BILOG version 3.0 program.

Table 2. Goodness of fit test of model by *p-value*

Category	Model		
	1PL	2PL	3PL
Fit	15	32	31
Unfit	25	8	9

Table 2 shows that the item analysis based on the Item Response Theory fits the 2PL model. The result of question analysis based on 2PL model in Package A Test found 27 good questions and 13 poor questions. Such poor questions were caused by the difficulty level and discrimination index that exceeded the criteria.

Table 3. Goodness of fit test of model by *p-value*

Category	Model		
	1PL	2PL	3PL
Fit	11	30	28
Unfit	29	10	12

Table 3 shows that the item analysis based on IRT fits the 2PL model. The result of the item analysis based on 2PL model in Package B Test found 24 good questions and 16 poor questions.

Information Function (IF). The item information function helps determining the quality of a test instrument. To observe the information function of Package A and B tests, 2PL model was used. In the 2PL model, the high-

est plot information function will be reached when a student who responds to an item has an ability that is equivalent to the difficulty level and discrimination index of the item.

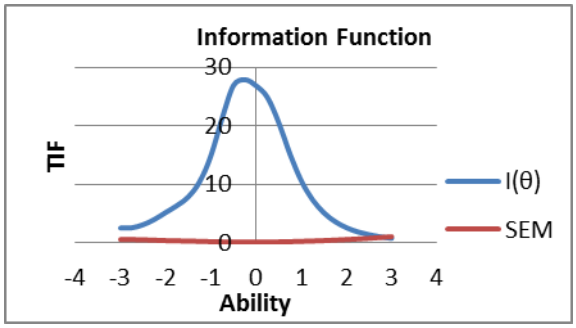


Figure 9. Chart of function information of Package A

Figure 9 shows that the maximum information function value is 27.884 with -0.250 logit (theta). The Estimated Standard Error of Measurement for Package A is 0.189 or inversely proportional with the information function of the test. This means that the participants of Accounting Vocational Theory trial Package A Test will give good information with the smallest measurement error if answered by the participants with -0.250 ability.

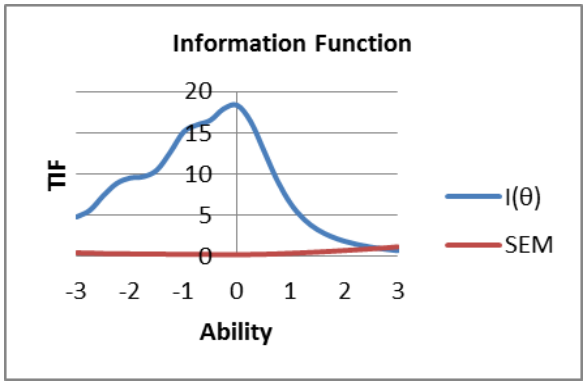


Figure 10. Chart of function information of Package B

Figure 10 indicates that the maximum information function value at 18.362 is reached with 0 logit (theta). The test's SEM is 0.2337 or inversely proportional with the test function. This means that the participants of Accounting Vocational Theory trial Package B Test will give good information with the smallest measurement error if the test was done by the participants with zero (0) ability.

Accounting Vocational Theory Trial Exam Equating Test. Verification of the equation of the Accounting Vocational Theory trial test of both Package A and B must be held in order to see whether or not both packages are parallel. The test for the test instruments' equation can be done using the t-test. The result of the t-test shows the significance value at equal variances assumed at 0.000 < alpha 0.05. This means that the average score in Package A and B differs (with the average difference of 3.092), and therefore, equating is necessary.

Equating

When the Accounting Vocational Theory trial exam instruments were proven unparallel, equating was necessary. During equating test, one needs to determine which package will be used as the benchmark. This research equated Package A to Package B, as presented in Table 4. Based on the result of analysis using BILOG 3.0, it is found that the items with good characteristics and the mostly fit are in the 2PL model.

Mean/Sigma Method. In the mean/sigma method, the calculation of α and β constants using the mean and standard deviation of the difficulty level resulted in constants α = 1.168 and β = 0.270. From the constants α and β, it is found the equation of Package A (*x*) to Package B (*y*) as follows:

$$\theta_2^* = 1.168\theta x + 0.270$$
$$b_2^* = 1.168bx + 0.270$$
$$a_2^* = \frac{a_1}{1.168}$$

Using α and β, item parameter transformation was carried out, which resulted in the equating item parameter as presented in Table 5. The Package A Test shows that there are 17 test items whose average difficulty level is -0.113 and standard deviation 0.641, and after equation, the mean changes to 0.138 and standard deviation changes to 0.749. Further, the average discrimination index of Package A test is 1.285 with the standard deviation of 0.386, and after equation the mean changes to 1.100, and the standard deviation changes to 0.330.

Table 4. Summary of question parameter

No	Package A		Package B	
	The difficulty level	Discrimination index	The difficulty level	Discrimination index
5	-0.320	1.364	0.843	1.084
8	-0.608	1.444	-0.677	1.719
9	-0.136	1.842	-0.282	1.793
12	-0.369	1.634	-0.949	1.606
13	-0.484	1.548	-0.307	1.709
16	-0.262	1.197	-0.154	1.167
17	-0.743	1.508	0.927	0.628
20	0.297	1.199	0.066	1.165
21	1.511	0.573	1.529	0.630
23	-0.103	1.552	0.169	1.787
24	-1.116	0.606	-0.270	1.848
27	-0.035	1.591	0.229	1.619
30	0.624	0.981	0.682	1.049
33	0.602	1.272	0.738	0.910
34	0.427	1.601	0.791	1.402
36	-0.468	1.317	0.391	0.887
37	-0.730	0.611	-1.375	0.905
μ	**-0.113**	**1.285**	**0.138**	**1.289**
σ	**0.641**	**0.386**	**0.749**	**0.422**

Table 5. Conversion of Package A to Package B using mean/sigma method

No	Package A		Package B	
	b Initial	aInitial	(b_2^*)	(a_2^*)
5	-0.320	1.364	-0.104	1.168
8	-0.608	1.444	-0.440	1.236
9	-0.136	1.842	0.111	1.577
12	-0.369	1.634	-0.161	1.399
13	-0.484	1.548	-0.295	1.325
16	-0.262	1.197	-0.036	1.025
17	-0.743	1.508	-0.598	1.291
20	0.297	1.199	0.617	1.026
21	1.511	0.573	2.035	0.491
23	-0.103	1.552	0.150	1.329
24	-1.116	0.606	-1.033	0.519
27	-0.035	1.591	0.229	1.362
30	0.624	0.981	0.999	0.840
33	0.602	1.272	0.973	1.089
34	0.427	1.601	0.769	1.371
36	-0.468	1.317	-0.277	1.128
37	-0.730	0.611	-0.583	0.523
μ	**-0.113**	**1.285**	**0.138**	**1.100**
Σ	**0.641**	**0.386**	**0.749**	**0.330**

Mean/Mean Method. In mean/mean method, the calculation of constants α and β uses the mean of difficulty level and discrimination index, which resulted in constants α = 0.997 and β = 0.250. From the constants α and β, it is found that the equation of Package A (*x*) to Package B (*y*) is as follows:

$$\theta_2^* = 0.997\theta x - 0.250$$
$$b_2^* = 0.997bx - 0.250$$
$$a_2^* = \frac{a_1}{0.997}$$

Table 6 shows the conversion of the result of equation to the difficulty level and discrimination index parameters. Package A test shows that there are 17 test items whose average difficulty level is -0.112 and standard deviation is 0.641, and after equation, the mean changes to 0.138 and standard deviation changes to 0.639. The parameter of discrimination index of Package A is 1.285 with the standard deviation of 0.385, and after equation, the mean changes to 1.289 and standard deviation changes to 0.387.

Table 6. Conversion of Package A to Package B using mean/mean method

No	Package A		Package B	
	b Initial	a Initial	(b_2^*)	(a_2^*)
5	-0.320	1.364	-0.069	1.368
8	-0.608	1.444	-0.356	1.448
9	-0.136	1.842	0.114	1.847
12	-0.369	1.634	-0.118	1.639
13	-0.484	1.548	-0.232	1.553
16	-0.262	1.197	-0.011	1.201
17	-0.743	1.508	-0.491	1.512
20	0.297	1.199	0.546	1.203
21	1.511	0.573	1.756	0.575
23	-0.103	1.552	0.147	1.557
24	-1.116	0.606	-0.863	0.608
27	-0.035	1.591	0.215	1.596
30	0.624	0.981	0.872	0.984
33	0.602	1.272	0.850	1.276
34	0.427	1.601	0.676	1.606
36	-0.468	1.317	-0.217	1.321
37	-0.730	0.611	-0.478	0.613
μ	**-0.112**	**1.285**	**0.138**	**1.289**
σ	**0.641**	**0.385**	**0.639**	**0.387**

Accuracy of Equating Result Based on Root Mean Square Difference. Kim and Cohen (1996, p. 17) explain the formula to calculate the equating accuracy as follows.

$$\text{RMSD}(a) = \sqrt{\frac{\sum_{i=1}^{N}(a_2^* - a_1)^2}{N}}$$

$$\text{RMSD}(b) = \sqrt{\frac{\sum_{i=1}^{N}(b_2^* - b_1)^2}{N}}$$

$$\text{RMSD}(\theta) = \sqrt{\frac{\sum_{i=1}^{N}(\theta_2^* - \theta_1)^2}{N}}$$

Note:
RMSD = Root Mean Square Difference
a_2^* = Differentiator power of the first test after being equated to the second test
a_1 = Differentiator power of the first test
b_2^* = The difficulty level of the first test after being equated to the second test
b_1 = The difficulty level of of the first test
θ_2^* = the ability of the test participants of the first test after being equated to the second test
θ_1 = the ability of the test participants of the first test

Table 7. Summary of RMSD calculation result for mean/sigma and mean/mean methods

Parameter	RMSD	
	Mean/Sigma Method	Mean/Mean Method
The difficulty level (b)	0.272	0.251
Discrimination index (a)	0.192	0.004
Ability (θ)	0.320	0.250

Table 7 shows that the RMSD value in the mean/mean method is lower than that of the RMSD value in mean/sigma method. It can be assumed that equation with the mean/mean method is more accurate compared to that with the mean/sigma method.

Discussion

Characteristics of Trial Exam Question Item Based on Question Item Review

Both Package A and B tests in the material aspect have 3 test items that require revision as they do not fit the exam content

outline. For the construction and language aspects, both Package A and B are 100% in good criteria.

Characteristics of Test items

The result of the analysis of Package A test shows that 15 test items fit the 1PL model, 32 test items fit the 2PL model, while 31 test items fit the 3PL model. The characteristics of questions in Package A based on the 2PL model show that there are 27 good questions that fit the model. Thirteen items are poor as their difficulty level and discrimination index do not meet the criteria (above +2).

The result of the analysis of Package B Test shows that 11 test items fit the 1PL model, 30 test items fit the 2PL model, and 28 test items fit the 3PL model. This shows that the 2PL model has the largest number of fit test items. If seen based on the 2PL model, 24 items are good and fit the model, while 16 items are poor.

Trial Exam Question Equating

The questions used in the trial exam of the Accounting Vocational Theory in Sleman Regency were given in Packages A and B. If both packages were used unequally, one of the student groups would be disadvantaged, particularly for students working on harder test packages. The result of the *t*-test on the scores in the two packages shows that both packages are non-parallel, and therefore, equating is necessary. The result of the question equating using the Mean/Sigma method resulted in the equation b_2^*= 1.168*bx* + 0.60, while the Mean/Mean method resulted in the equation b_2^* = 0.997*bx*– 0.250.

Kilmen and Demirtasli (2012) conduct similar equating research by using four methods in the IRT approach. Those four methods use the least RMSD value to determine the accuracy. The RMSD value in the mean/mean method is smaller than the RMSD value in the mean/sigma method. The mean/mean method resulted in the RMSD for parameter b at 0.251, parameter a at 0.004, and ability parameter at 0.250 whereas the mean/sigma method resulted in the RMSD for parameter b at 0.272, parameter a at 0.192, and ability parameter at 0.320. The lower RMSD value shows more accurate equating result, in this case, it is shown that the mean/mean equating method shows better result than the mean/sigma method.

Conclusion

The results of expert review of the test items are as follows. (1) In terms of the material, construction, and language aspects, the test items in the test instruments of Accounting Vocational Theory trial exam prepared by Accounting Subject-Matter Teachers' Forum of Sleman Regency are in a good category. (2) The content validity of Package A and B test items according Aiken formula is satisfactory. (3) The reliability coefficient of test instruments of Accounting Vocational Theory trial exam for both Package A and B is in a good category, at 0.887 for Package A and 0.856 for Package B. (4) The analysis based on the Item Response Theory using the 2PL model to Package A test shows that 32 items fit the model, whereas 30 items fit the model of Package B. The discrimination index of Package A shows that there are 27 good items and 13 poor items. In Package B test, 24 items are in a good category while the remaining 16 items are in a poor category. Poor items are resulted from the difficulty level and discrimination index which exceed the criteria. (5) Equation using the Mean/Mean method shows smaller result compared to the RMSD value found using the Mean/Sigma method.

References

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

Aminah, N. S. (2012). Karakteristik metode penyetaraan skor tes untuk data dikotomos. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *16*(Special Issue for UNY's 48th Dies-Natalis), 88–101. https://doi.org/10.21831/pep.v16i0.1107

Cook, L. L., & Eignor, D. R. (1991). An NCMF instructional module on IRT equating methods. *Educational Measurement: Issues and Practice*, *10*, 37–45.

Crocker, L. M., & Algina, J. (2008). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart, and Winston.

Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). *Principles and practices of test score equating*. Princeton, NJ: Educational Testing Service.

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

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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage Publications.

Kilmen, S., & Demirtasli, N. (2012). Comparison of test equating methods based on Item Response Theory according to the sample size and ability distribution. *Procedia - Social and Behavioral Sciences*, *46*, 130–134. https://doi.org/10.1016/J.SBSPRO.2012.05.081

Kim, S.-H., & Cohen, A. S. (1996). A comparison of linking and concurrent calibration under Item Response Theory. In *American Educational Research Association Annual Meeting* (pp. 1–52). New York, NY: American Educational Research Association.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. New York, NY: Springer.

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

Liao, C.-W., & Livingston, S. A. (2012). A search for alternatives to common-item equating. In *paper presented at the annual meeting of the National Council on Measurement in Education*. Vancouver, British Columbia, Canada.

Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.

Miyatun, E., & Mardapi, D. (2000). Komparasi metode penyetaraan tes menurut teori respons butir. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *2*(3), 1–18. https://doi.org/10.21831/pep.v2i3.2083

Moghadamzadeh, A., Salehi, K., & Khodaie, E. (2011). A comparison method of equating classic and Item Response Theory (IRT): A case of Iranian study in the university entrance exam. *Procedia - Social and Behavioral Sciences*, *29*, 1368–1372. https://doi.org/10.1016/j.sbspro.2011.11.375

Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston, MA: Pearson Education.

Regulation of the Minister of Education No. 20 of 2007 on the educational assessment standard (2007). Republic of Indonesia.

Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.

Sukirno, S. (2007). Penyetaraan Tes UAN: Mengapa dan Bagaimana? *Cakrawala Pendidikan*, *26*(3), 305–321. https://doi.org/10.21831/cp.v3i3.3983

REiD (Research and Evaluation in Education), 3(2), 2017, 163-173
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

The utilization of junior high school mathematics national examination data: A conceptual error diagnosis

***[1]Kartianom; [2]Djemari Mardapi**
*Graduate School of Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: kartianom@gmail.com

Submitted: 23 January 2018 | Revised: 22 February 2018 | Accepted: 26 February 2018

Abstract

The goal of the research is to gain insights into the characteristics of the items in the mathematics national examination, the attributes on which the items were formulated and the result of a conceptual error diagnosis of the mathematics materials based on the result of the junior high school mathematics national examination. This is quantitative descriptive research. The data were collected from 3,079 grade-nine students of junior high schools who took the National Examination in the academic year of 2015/2016. The sample was established randomly based on the package code of the examination which is P0C5520 with 574 students as the examinees. Documentation method was applied in collecting the data. The result of the research shows that – upon the implementation of the classical test theory – there are 16 items in 'difficult' category, 24 in 'intermediate' category, and no items in 'easy' category. Furthermore, upon the implementtation of the item response theory, the result shows that 28 items are in 'good' category and 12 items are in 'poor' category. In addition, there are 50 attributes on which the Junior High School Mathematics National Examination test (package P0C520) is formulated. Four attributes are content attributes and the rest (46) are process skill attributes. The result of the diagnosis shows that there are 11 types of errors made by the students when trying to complete the content items. Most of the errors are conceptual errors related to the geometric materials especially in the submaterials of polyhedron, triangles, and quadrangles.

Keywords: *conceptual error, attributes, junior high school mathematics national examination*

How to cite item:

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. doi:http://dx.doi.org/10.21831/reid.v3i2.18120

Introduction

In the education system, evaluation is an urgent thing to perform. Evaluation is a medium to put students in the context of what they understand and what they are able to perform, while describing what they do not understand and what they are not able to perform (Sumintono & Widhiarso, 2015, pp. 2–3). The goal of the evaluation on the result of the study as conducted by the government is to measure the competence level of the graduates on certain subjects as formulated in National Examination (or *Ujian Nasional* – UN). The items in National Examination are formulated based on the competence standards of the graduates, basic competence and achievement indicator.

Most of the education practitioners utilize the reports on the result of the National Examination as the supporting data in the process of policy-making, as a medium in

comparing the achievement of the examinees in the national level and as a medium in mapping the quality of national education. For example, the report of the Junior High School National Examination result for Mathematics in Baubau Municipality in the academic year of 2014/2015 shows that the average score on Mathematics is $(\bar{X})$ 42.62 with 15.0 as the lowest score and 97.5 as the highest score (Ministry of Education and Culture, 2015). The result indicates that some examinees gave incorrect responses to some of the items of the Mathematics National Examination. The mistakes might be caused by the level of the items in the examination and the examinees' lack of conceptual knowledge or because they made a conceptual errors.

A good examination item must go through a calibration process, so the information on the items can be gained from the applied test. This information is commonly called characteristics of the items, which can be estimated by using two approaches, namely: Classical Test Theory (CTT) and Item Response Theory (IRT). A good item can be reviewed from its difficulty level, discrimination index, and distractor effectiveness. In the CTT approach, the index of the difficulty level of a good item must be 0.3 – 0.8, while the discrimination index must be ≥ 0.3 and the option of each item at least has to be selected by 5% of the examinees (Mardapi, 2012, p. 128). In the IRT approach, the index of the difficulty level of a good item must be (*ai*) -2.0 – +2.0 (Hambleton, Swaminathan, & Rogers, 1991, p. 13), while the discrimination index must be (*bi*) 0 - +2.0 (Hambleton et al., 1991, p. 15), and pseudo guessing index must be (*ci*) 0 – 1/k (Hambleton et al., 1991, p. 17).

Items with very low or very high facility index cannot be categorized as good items because they cannot differentiate the level of ability of the examinees. The error indication of the examinees can be caused by the difficulty level. It might not be caused by the lack of competence. Items with negative discrimination index indicate that the correctness of the answer is questionable. The correctness of the answer is also questionable if the distracting items are only selected by <5% of the examinees. The examinees with the pseudo guessing index >1/k show that the distracting items are not able to attract those with low capability (Abadyo & Bastari, 2015).

A conceptual error is an error in understanding the concept in which the understanding is not in accordance with the scientific definition as agreed generally by the experts in that field. In mathematics, this error happens when students fail to relate the initial concept with the newly-given one (Russell, O'Dwyer, & Miranda, 2009, p. 416). In fact, a conceptual error is closely related to the conceptual knowledge of the examinees. Mathematics conceptual knowledge is the examinees' understanding of the scope of the field of mathematics. The scope of mathematics subject include: (1) number, (2) algebra, (3) geometry and measurement, and (4) statistics and probability. Therefore, in mathematics, a conceptual error can be defined as an incorrect use of the concepts which do not follow the scientific definition in the scope of mathematics field (numbers, algebra, geometry, and measurement and statistics and probability.

In order to learn about the error indication related to a conceptual error, there should be diagnosis process. The goal of the diagnosis activity is to understand the strength and weakness of the examinees (Leighton & Gierl, 2007, p. 242). The cognitive diagnosis model (CDMs) can be utilized in two ways, (a) retrofitting (post-hoc analysis) from non-diagnostic examination to gain richer or wider information and (b) designing or constructing a set of items for diagnostic purposes (Ravand & Robitzsch, 2015, p. 3). In the approach of retrofitting (post-hoc analysis), non-diagnostic examination instruments are reconstructed in a way that they can be used to identify the strength and weakness of the examinees in defining the attributes based on which the test items are formulated.

Attributes are the description of knowledge in completing examination contents in a certain domain (Wang & Gierl, 2011, p. 166) and the basis of cognitive or skill process crucial to completing the test items (Gierl, Cui, & Zhou, 2009, p. 5; Gierl, Zheng, & Cui, 2008, pp. 66–67; Yamtinah & Budiyono, 2015, p. 71). In mathematics, attributes consist of three categories: content attributes (common

materials), process attributes (expected capability after learning the materials in the content attributes) and skill attributes (specific mathematical skills critical in certain materials) (Tatsuoka, 2009, p. 2). Attributes utilized in this research are content attributes and process skill attributes.

There are already many studies taking advantages of diagnosis activities in Indonesia. However, most of them focus on the development of the diagnostic instruments. Secondary data such as national examination, PISA and TIMSS are rarely used in diagnostic activities. If we take a look at the studies in the last six years (2011-2017), secondary data have been a fresh medium to gain information on the influential factors in the academic achievement of examinees (Kartianom & Ndayizeye, 2017, p. 200) and the difficulty of the examinees in completing the mathematics test items of the National Examination (Isgiyanto, 2011, p. 308; Retnawati, 2017, p. 33). Even though National Examination is neither the main factor in determining the passing of the examinees, nor the main requirement in continuing to higher education level, the result of the National Examination is valuable data for diagnostic purposes.

To be more specific, the poor result of the Junior High School National Examination in Baubau Municipality was driven by the lack of comprehensive diagnosis on the result of the National Examination, especially on the subject of Mathematics. Both of the academia and the municipality administrator do not seem to see diagnostic activities as an urgent matter. The data of the National Examination are left untouched and have not yet been transformed into insightful information. The objective of this research is to gain insights into the characteristics of the test items and see the result of the diagnosis on the conceptual error in mathematics materials based on the result of the Junior High School Mathematics National Examination in Baubau Municipality.

Method

This research is quantitative descriptive research which applies content analysis in drawing conclusion by identifying various characteristics specifically in a message – in the test items and the responses of the examinees - objectively, systematically and generally. The research was conducted in Baubau Municipality. The data were collected from the Center for Education Evaluation (commonly known as PUSPENDIK) in Jakarta, in the form of National Examination sheets and the response sheets.

The data source is the ninth graders of junior high schools in the academic year of 2015/2016 in Baubau Municipality. The total number of the examinees is 3,079. The sample was established randomly (random sampling) based on the package code of the examination content. The researchers selected the package code of P0C5520 with 574 examinees in total. The object of the research is 40 test items and 22,960 responses of the examinees.

The *expost facto* data in the form of the the examinees' responses and the items in the Junior High School Mathematics National Examination were collected using documentation technique. The data were analyzed for diagnostic information. The items in the National Examination were selected to be the data because they had been standardized. Therefore, the bias has been minimized. Moreover, they had been calibrated, which allowed the researchers to compare the existing series and the packages from each year.

A good examination instrument must be valid and reliable. In this research, the instruments chosen are the instruments of the National Examination which have been tested in large and small scales. Therefore, it is safe to assume that the validity and reliability of the instruments are fulfilled. The validity implemented in this research is closely related to the attribute formation. The validity of the content of the attributes on which the test items are formulated was proven based on the judgment of the experts. In order to produce the content validity index of the attributes formation, the result of the judgment was then calculated using Aiken formulation. Based on the Aiken index, the researchers formulated criteria in order to show the content validity of the attributes formation (see Table 1) (Kartianom, 2017, p. 153).

Table 1. Content validity index criteria

Aiken Index	Content Validity Criteria
> 0.4	Low
0.4 – 0.8	Medium
> 0.8	High

In order to understand the characteristics of the items using CTT approach, the data were analyzed using TAP software version 14.7.4. Table 2 shows the criteria of good items based on CTT approach (Mardapi, 2012, p. 128).

Table 2. Item characteristic criteria using CTT

Parameter	Criteria
a_i	More than or equal with 0.3
b_i	0.3 to 0.8
c_i	The answer choice is chosen by at least 5% of the examinees

Description:
a_i = Items differentiators index
b_i = Items difficulty level index
c_i = Distractor effectiveness index

Using IRT approach, the data were analyzed with the help of Bilog-Mg software. Prior to the analysis, the sample was tested for its adequacy using SPSS11.5 software. The sample is considered adequate when the value of *Kaiser Mayer Olkin Measure* (KMO) > 0.5 with significance value (Sig.) of < 0.05. After that, the assumption test was conducted on the item parameter estimation using IRT approach. The assumption to be fulfilled was local unidimension and independency. Unidimension assumption was conducted with the support of SPSS 11.5 software based on the formation of the dominant factor. The formulated factor was with the Eigen value > 1.0. The dominant factor has large Eigen value discrepancy with the next factor and it has at least 20% cumulative frequency (Retnawati, Munadi, & Al-Zuhdy, 2015). The local independency assumption will be automatically fulfilled when the unidimensional assumption is fulfilled (Retnawati, 2014, p. 141).

When the assumption in IRT approach has been fulfilled, the next one is goodness of fit test. There are three models in IRT approach: model 1-PL, model 2-PL and model 3-PL. The goodness of fit test is conducted with the support from Bilog-Mg software by comparing the significant value of χ^2 with $\alpha = 0.05$ and also ICC curve. If the value of sig. $\chi^2 > \alpha = 0.05$, the items can be categorized as fit with the model. For ICC curve, the data are considered fit when the distribution of the data matches the model (Figure 1).

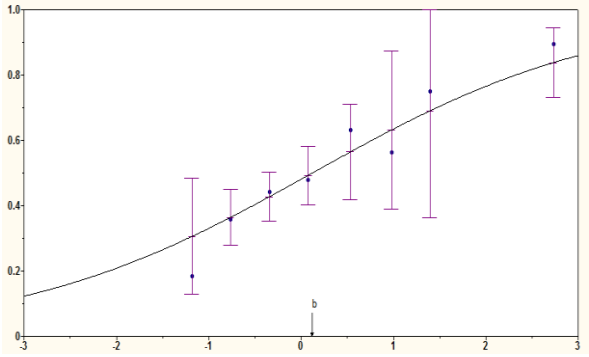


Figure 1. ICC curve

In each model, the criteria of good items in the IRT approach are presented in Table 3 (Hambleton et al., 1991, pp. 13–17).

Table 3. IRT criteria of items characteristics

Model	Parameter Criteria		
	a_i	b_i	c_i
1-PL	0 up to +2	-	-
2-PL	0 up to +2	-2 up to +2	-
3-PL	0 up to +2	-2 up to +2	0 up to 1/k

Description:
a_i = Item discrimination index
b_i = Items difficulty level index
c_i = Pseudo guessing index

In this research, the error made by the examinees was analyzed through the response of the Mathematics examination contents (answer sheets of the examinees) of the National Examination in the academic year of 2015/ 2016. The analysis was conducted by formulating the probable description of the alternative response to the test items. At this point, the researchers did not use the description of the examinees' answers and the responses to determine the achievement of the students, but to understand the type and the area of the error.

In order to conduct the diagnosis on the a conceptual error made by the examinees, the researchers: (1) identified the attributes of the examination content by defining the op-

tions of responses to each item using the content analysis; (2) named the type of the error in each response option based on the attributes on which the items were formulated; (3) analyzed the response option using TAP software version 14.7.4 to measure the percentage of each type of error in each material. There was a follow up for the most dominant type of error in order to understand the area of the error.

Findings and Discussion

The Characteristics of the Test Items

Classical Test Theory

To understand the difficulty level, differentiator, and distractor effectiveness of the examination content, the researchers applied the classical test theory when analyzing the items. The data were in the form of answer sheets - multiple choices with the answer key. Table 4 shows the result of the recapitulation of the characteristics of the test items based on the difficulty level of the items in each material.

Table 4. The difficulty level of the items in each material

Materials	Category			Total
	Easy	Medium	Difficult	
Numbers	0	7	4	11
Algebra	0	4	6	10
Geometry	0	9	4	13
Statistics	0	3	1	4
Probability	0	1	1	2
Total	**0**	**24**	**16**	**40**

Table 4 shows that: (1) the materials on number have seven items in 'medium' category and four items in 'difficult' category; (2) the materials on algebra have four items in 'medium' category and six items in 'difficult' category; (3) the materials on geometry have nine items in 'medium' category and four items in 'difficult' category; (4) the materials on statistics have three items in 'medium' category and one item in 'difficult' category; and (5) the materials on probability have one item in 'medium' category and one item in 'difficult' category.

Table 5 shows the result of the recapitulation of the characteristics of the test items based on the differentiators of the items in each material.

Table 5. The differentiators of the items in each materials

Materials	Category		Total
	Good	Not Good	
Numbers	9	2	11
Algebra	6	4	10
Geometry	8	5	13
Statistics	1	3	4
Probability	2	0	2
Total	**26**	**14**	**40**

Table 5 shows that overall the discrimination index of the test items in the content of the Mathematics National Examination in Baubau Municipality has 26 items in 'good' category and 14 items in 'not good' category. If we take a closer look at the materials: (1) the materials on numbers have nine items in 'good' category and two items in 'not good' category, (2) the materials on algebra have six items in 'good' category and four items in 'not good' category, (3) the materials on geometry have eight items in 'good' category and five items in 'not good' category; (4) the materials on statistics have one item in 'good' category and three items in 'not good' category; and (5) the materials on probability have two items in 'good' category and no item is in 'not good' category.

Other critical information in the classical test theory is distractors effectiveness. The distribution of the response choice can be considered as effective or acceptable when each option in the test items is chosen by at least 5% of the examinees (Mardapi, 2012, p. 129). Figure 2 presents the functionality percentage of the distracting items.

Figure 2. The functionality percentage of the distractors

Figure 2 shows that 100% of the items have effective distractors. This means the distractors in the items of the Junior High School Mathematics National Examination in Baubau Municipality are well-functioned distractors. In other words, they are able to attract the examinees.

Item Response Theory

Principally, the item response theory uses the probabilistic model. There are three analytic models: 1PL, 2PL and 3PL. In order to correctly select analytic model, the goodness of fit test is a crucial process. However, before that, the sample adequacy and assumption test has to be conducted. Table 6 shows the result of the sample adequacy test.

Table 6. The result of the *KMO* and *Bartlett*
KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.810
Bartlett's Test of Sphericity	Approx. Chi-Square	2425.233
	df	780
	Sig.	0.000

Table 6 shows that the KMO value is at 0.810 or 0.5 higher. This means that the sample used in this research is adequate. Next, unidimensional assumption test was conducted while considering the scree plot (Figure 3).

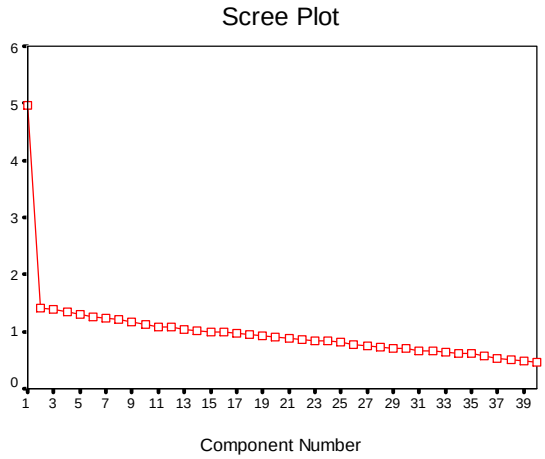


Figure 3. The scree plot of the result of the exploratory factor analysis

The scree plot in Figure 3 shows that there is one dominant factor in the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality. This can be seen from the shift in the Eigen value of the first factor up to the second factor. In the second factor and beyond, the shift of the Eigen value is not too high. Therefore, it is safe to conclude that the unidimensional assumption test on the contents of the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality has been fulfilled. When the unidimensional assumption test has been fulfilled, the local independency assumption is automatically fulfilled. This also means that there is a correlation among the factors in the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality, so the goodness of fit test can be conducted. The goodness of fit test for models 1-PL, 2-PL and 3-PL is conducted by comparing the significant value of χ^2 with $\alpha = 0.05$ and ICC curve. Table 7 shows the result of the goodness of fit test for 1-PL, 2-PL and 3-PL.

Table 7. The result of the goodness of fit between the items and the model

Fitting Model	**Fitting Items**		
	Model 1-PL	**Model 2-PL**	**Model 3-PL**
Sig. *Chi-Square* Value	24	35	13
Using ICC curve	5	12	2

Table 7 shows that based on the goodness of fit test, 24 items fit with model 1-PL, 35 items fit with model 2-PL and 13 items fit with model 3-PL. When the goodness of fit test with ICC curve is applied, five items fit with model 1-PL, 12 items fit with model 2-PL and two items fit with model 3-PL. This makes model 2-PL the fittest analytic model.

The parameter used in model 2-PL is the difficulty level (*bi*) and differentiators (*ai*), whereas guessing (*ci*) for the item is considered zero. The items which fit with model 2-PL are brought to the next analytic step. The items are as follows, items 1, 2. 3, 4, 5, 7, 8, 9, 10. 12. 13, 14, 15, 16, 17, 19, 20. 21, 22. 24, 25, 26, 27, 29, 30. 31, 32. 33, 34, 35, 36, 37, 38, 39 and 40. In model 2-PL, the items that do not fit with model 2-PL are not included

in the next analytic steps even though they have difficulty and differentiators as the parameter. These excluded items are items 6, 11, 18, 23 and 28.

Table 8 shows the result of the characteristics analysis on the test items based on model 2-PL with the support from Bilog-MG program.

Table 8. The characteristics of the test items based on the parameter of difficulty level and differentiators

Category	Parameter Frequency		Desc.
	a	b	
Good	35	28	28
Not Good	0	7	7
Total	**35**	**35**	**35**

Table 8 shows that based on the criteria of model 2-PL, there are 28 items in 'good' category and 7 items in 'not good' category. In fact, those 7 items in 'not good' category possess good differentiators but have bad difficulty level. Those items are items 33, 9, 15, 29, 19, 21, and 35. Respectively, their difficulty level parameters are 4.463, 4.027, 3.870, 2.747, 2.644, 2.100, and 2.028. These items have very high difficulty level with item 33 having the highest difficulty level. In terms of the differentiator's parameter, 40 items fall in 'good' category. This strengthens the indication that the error in the examinees responses – specifically while trying to complete items 33, 9, 15, 29, 19, 21 and 35 – is not caused by the difficulty level. In addition to items parameter, the researchers also gain insights into the test information function as shown in Figure 4.

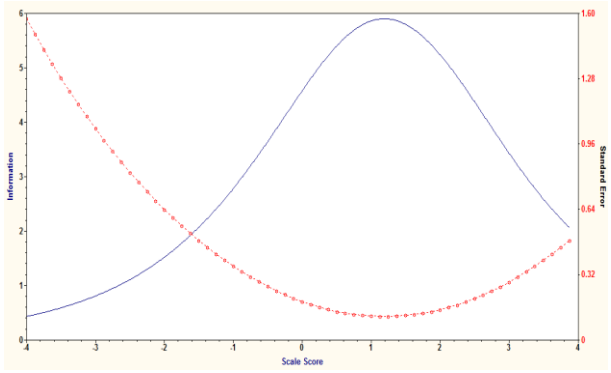


Figure 4. Information functions and test measurement error

Figure 4 shows that the content of Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality has higher information than the error in measurement with the ability range from -1.6 to +4.0. If the examination was delivered to the examinees with the ability range lower than -1.6 and higher than +4.0, the error in the measurement would be a lot higher than the information function.

Subject-Matter Mastery in the Mathematics National Examination

The subject-matter mastery of the test takers of the National Examination of Mathematics of the academic year 2015/2016 can be seen from the proportion of true answers of the test takers on the number, algebra, geometry, statistics, and probability materials as presented in Figure 5.

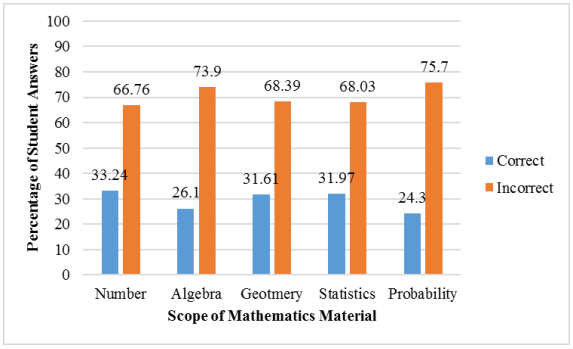


Figure 5. Percentage of student's answers to each material

Figure 5 shows that all materials tested on the Mathematics National Examination of the academic year 2015/2016 in Baubau Municipality are considered difficult by the test takers. This can be seen from the percentage of the wrong answers that are greater than the percentage of the correct answers of the test takers on each material.

Attributes on which Test Items are Formulated

The attributes, on which the items are formulated, are developed and validated by five experts (expert judgment), three of whom are mathematics teachers of state junior high schools in Yogyakarta who previously had in-

volved in the development of the examination, and two are Mathematics lecturers. Generally, all of the attributes of the items of the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality consist of four content attributes and 46 process skill attributes. The content validity index of the attributes of those 40 items is at 0.888 which falls in 'high' category. Table 9 shows the distribution of the attributes of the items in each material.

Table 9. The distribution of the test items attributes

No	Material	Content Attributes	Process Skill Attributes
1	Numbers	1	13
2	Algebra	1	13
3	Geometry	1	14
4	Statistics and Probability	1	6
	Total	4	46

Table 9 shows the distribution of the attributes on which the test items are formulated. Each material competence has several attributes. Some of the attributes are alike and some are different. Thus, the material competence has to be divided into groups along with all of the attributes.

Diagnosis of the Examinees' Errors

Error Type

The identification of the error focuses on the attributes which are not mastered and applied correctly by the examinees when they are trying to complete the items in the Mathematics National Examination. Based on the content analysis, the errors can be categorized into 11 types, which consist of: (1) conceptual errors, (2) language-related interpretative errors, (3) procedural errors, (4) calculation errors, (5) representation errors, (6) conceptual and language-related interpretative errors, (7) conceptual and calculation errors, (8) conceptual and calculation errors, (9) language-related interpretative and procedural errors, (10) representation and procedural errors, and (11) representation and calculation errors. Figure 5 shows the percentage of each type of error.

Furthermore, in general, Table 10 shows the frequency of each type of errors. Table 10 shows that most of the errors are conceptual errors. They are in the area of basic concept of numbers, algebra, geometry (plane figure and solid figure) and probability. Most of them are found in geometric materials.

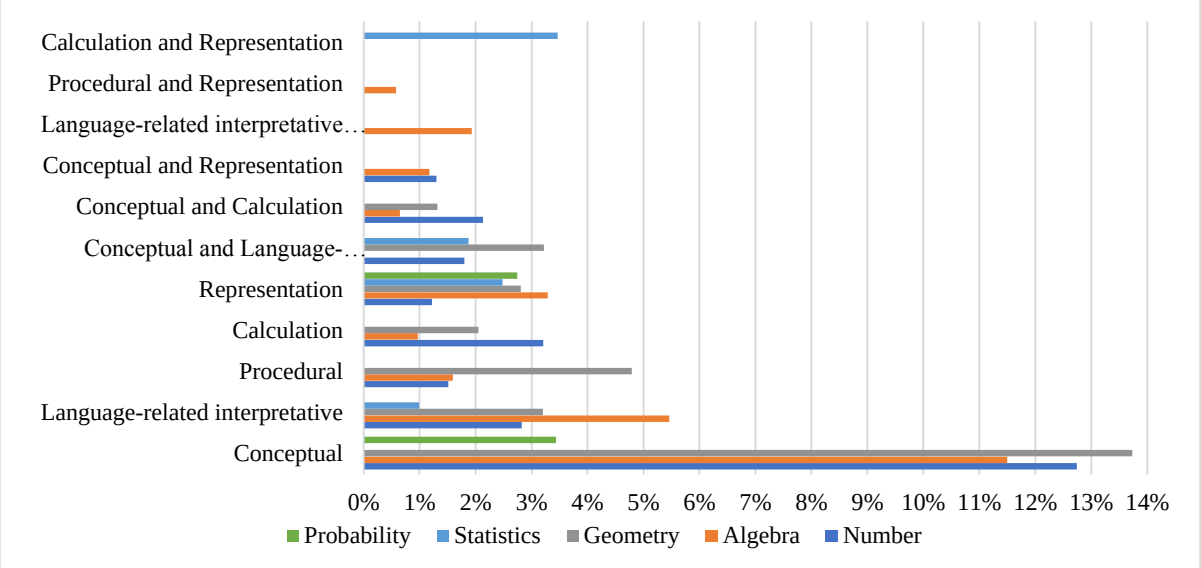


Figure 5. The percentage of each type of error in each material

Table 10. Types of errors made by the examinees

Types of Errors	Frequency	Percentage (%)
Conceptual	5804	41.41
Language-related interpretative	1749	12.48
Procedural	1106	7.89
Calculation	873	6.23
Representation	1759	12.55
Conceptual and Language-related interpretative	966	6.89
Conceptual and Calculation	575	4.10
Conceptual and Representation	347	2.48
Language-related interpretative and Procedural	271	1.93
Procedural and Representation	81	0.58
Calculation and Representation	486	3.47
Total	**14017**	**100**

The Area of the Conceptual Errors

The most dominant conceptual errors are: (1) the basic concept of integers in the materials of numbers, root form (irrational) and comparison; (2) the concept of relation and function, basic concept of algebraic operation, basic concept of integers and straight line equation in the materials of algebra; (3) the basic concept of geometry, polyhedron, triangles and quadrangles in the materials of geometry; (4) the basic concept of probability in the materials of statistics. These all are shown in details in Figure 6.

Discussion

By using CTT and IRT, there are five items with a very high level of difficulty (Items 9,15,19,21, and 33). Item 9 is related to number; items 9, 15 and 21 are about algebra, while item 33 is related to geometry. The high percentage of students answering those items wrongly is due the very high level of item difficulty. Besides, the very high level of item difficulty indicates that there are a lot of students with incomplete attributes of those materials.

Based on the content analysis, there are 11 types of students' errors. The conceptual error is the dominant type of errors mostly occured in geomerty-related items. In line with the result of this research, Isgiyanto (2011) also found that, in Indonesia, the junior high school students are weak at geometry and measurement with the low level of attributes of content/concept completeness.

The conceptual errors made by the students are indicated by the conceptual errors occurring in number and algebra materials. The testees' understanding of numbers is the key to understanding the material of algebra. The understanding of numbers and algebra is the requirement for the understanding of the geometrical materials. Further, in their study,

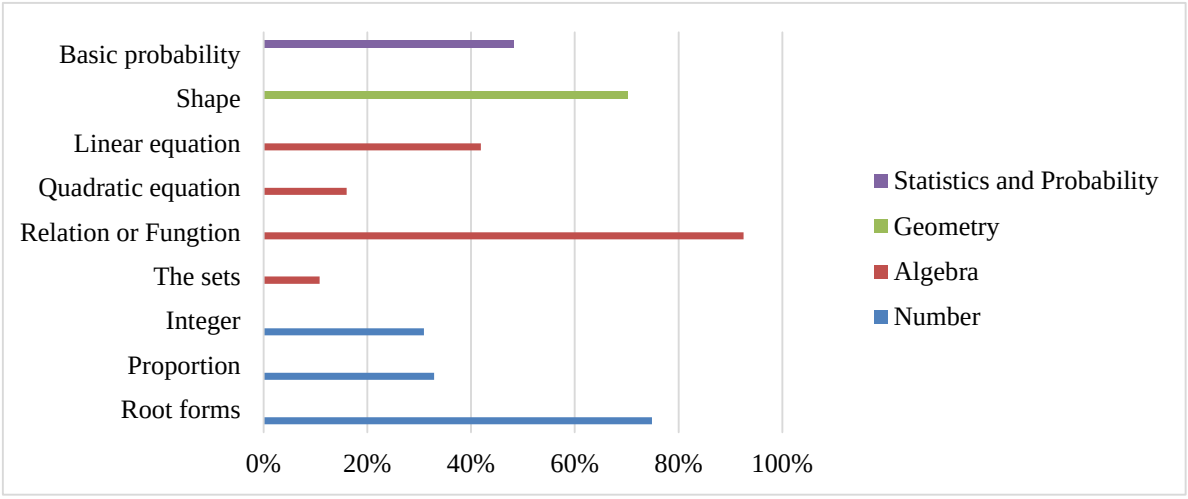


Figure 6. The area of error in each material

Russell et al. (2009, p. 416) mention that a conceptual error occurs because of the failure in connecting new concept with the earlier concept. Specifically, the conceptual error made by the students is located in the basic concept of integers, irrationals, comparisons, association and function, algebra operation, linear equation, polyhedron geometry, triangle, square, and probability.

The findings of this research are supported by the findings of a research conducted by Retnawati (2017, p. 33), which found that junior high school students in Yogyakarta, Indonesia found it difficult to finish the National Examination questions due to their disability to understand the concept of fraction, rationing fraction with square-root denominator, linear equation with one or two variables, determining the members of a sets, determining the gradient a linear equation, also the concept of area.

Conclusion and Recommendations

Conclusion

Based on the result of the analysis and description, it can be concluded that, *first*, based on the classical test theory, 16 test items are in 'difficult' category, 24 are in 'medium' category, and no item is in 'easy' category. Based on item response theory, 28 items are in 'good' category and 12 items are in 'not good' category. *Second,* there are 50 attributes – 4 content attributes and 46 process skill attributes - on which the Junior High School Mathematics National Examination content (package P0C5520) are formulated. *Third*, there are 11 types of errors made by the examinees when they tried to complete the examination. Most of the errors are conceptual errors in the materials of geometry especially in the sub materials of polyhedron, triangles and quadrangles.

Recommendation

Based on the conclusion, the recommendations are: (1) for users of the diagnostic information. The result of the research can be used as the materials for training on the process of conducting diagnostic information. It is expected that this type of training can be used to improve the quality of learning process in the schools with low result in the Mathematics National Examination. (2) For researchers, this research focuses only on diagnosis the types and areas of error made by the examinees when trying to complete Junior High School Mathematics National Test items based on the attributes of the items. Therefore, this research can be deepened by diagnosing the errors or difficulties faced by the examinees with the help of R packages CDM program while using model DINA.

References

Abadyo, A., & Bastari, B. (2015). Estimation of ability and item parameters in mathematics testing by using the combination of 3PLM/GRM and MCM/GPCM scoring model. *REiD (Research and Evaluation in Education)*, *1*(1), 55–72.

Gierl, M. J., Cui, Y., & Zhou, J. (2009). Reliability and attribute-based scoring in cognitive diagnostic assessment. *Journal of Educational Measurement*, *46*(3), 293–313. https://doi.org/10.1111/j.1745-3984.2009.00082.x

Gierl, M. J., Zheng, Y., & Cui, Y. (2008). Using the attribute hierarchy method to identify and interpret cognitive skills that produce group differences. *Journal of Educational Measurement Spring*, *45*(1), 65–89. Retrieved from https://pdfs.semanticscholar.org/0a0b/180342ee51f6121dd4e3199c9cc4df3bc377.pdf

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. New Delhi: Sage Publications.

Isgiyanto, A. (2011). Diagnosis kesalahan siswa berbasis penskoran politomus model partial credit pada matematika. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *15*(2), 308–325. Retrieved from https://journal.uny.ac.id/index.php/jpep/article/view/1099/1151

Kartianom, K. (2017). *Diagnosis kesalahan konsep materi matematika SMP berdasarkan hasil ujian nasional di kota Baubau*. Master Thesis, Universitas Negeri Yogyakarta,

Indonesia.

Kartianom, K., & Ndayizeye, O. (2017). What 's wrong with the Asian and African students' mathematics learning achievement? The multilevel PISA 2015 data analysis for Indonesia, Japan, and Algeria. *Jurnal Riset Pendidikan Matematika*, *4*(2), 200–210. https://doi.org/10.21831/jrpm.v4i2.16931

Leighton, J. P., & Gierl, M. J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, *26*(2), 3–16. https://doi.org/10.1111/j.1745-3992.2007.00090.x

Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.

Ministry of Education and Culture. (2015). *Laporan hasil ujian nasional*. Jakarta: Balitbang.

Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation*, *20*(11). Retrieved from http://pareonline.net/getvn.asp?v=20&n=11

Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.

Retnawati, H. (2017). Diagnosing the junior high school students' difficulties in learning mathematics. *International Journal on New Trends in Education and Their Implications*, *8*(1), 33–50. Retrieved from http://www.ijonte.org/FileUpload/ks63207/File/04.heri_retnawati.pdf

Retnawati, H., Munadi, S., & Al-Zuhdy, Y. A. (2015). Factor analysis to identify the dimension of Test of English Proficiency (TOEP) in the listening section. *REiD (Research and Evaluation in Education)*, *1*(1), 45–54. https://doi.org/10.21831/reid.v1i1.4897

Russell, M., O'Dwyer, L. M., & Miranda, H. (2009). Diagnosing students' misconceptions in algebra: Results from an experimental pilot study. *Behavior Research Methods*, *41*(2), 414–424. https://doi.org/10.3758/BRM.41.2.414

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan*. Bandung: Trim Komunikata.

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York, NY: Routledge/Taylor & Francis.

Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, *48*(2), 165–187. https://doi.org/10.1111/j.1745-3984.2011.00142.x

Yamtinah, S., & Budiyono, B. (2015). Pengembangan instrumen diagnosis kesulitan belajar pada pembelajaran kimia di SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *19*(1), 69–81. https://doi.org/10.21831/pep.v19i1.4557

REiD (Research and Evaluation in Education), 3(2), 2017, 174-182
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Quartet cards as the media of career exploration for lower-grade primary school students

***[1]Yulia Ayriza; [2]Farida Agus Setiawati; [3]Agus Triyanto; [4]Nanang Erma Gunawan; [5]Moh Khoerul Anwar; [6]Nugraheni Dwi Budiarti**
*Department of Psychology, Faculty of Education, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Karangmalang, Depok, Sleman, Yogyakarta 55281, Indonesia
Email: yulia_ayriza@uny.ac.id

Submitted: 16 January 2018 | Revised: 20 February 2018 | Accepted: 26 February 2018

Abstract

A career developed through the optimization of one's potentials will irrevocably play a role in the development of self-identity as well as the psychological well-being of the individual. When children are introduced and allowed to explore as many career options as possible during their developmental stage, they are more likely to have a fruitful career development in the future. The preceding study showed that the career interests and knowledge of lower-grade primary students fit the Holland Career Categories: realistic, investigative, artistic, social, enterprising, and conventional (RIASEC). It was also found that the students' career interest and knowledge levels varied, with most in the low level. This second-year study aims to expand the results of the previous study by developing the use of Quartet cards as the media of career exploration for lower-grade primary students. By using the research and development method, this study develops Quartet Career Cards into three difficulty levels: low, medium, and high. The Quartet cards media have undergone feasibility tests conducted by experts in theory and media, as well as a series of field testing consisting of preliminary, main, and operational stages among a total of 266 primary students of grades 1, 2, and 3. A revision was made on several components including the images, information, colors, font sizes, illustration styles, and card sizes. The findings show that Quartet Career Cards meet the feasibility standards for the media of career exploration.

Keywords: *career exploration, Quartet Career Cards, lower-grade primary school students*

How to cite item:

Ayriza, Y., Setyawati, F., Triyanto, A., Gunawan, N., Anwar, M., & Budiarti, N. (2018). Quartet cards as the media of career exploration for lower-grade primary students. *REiD (Research and Evaluation in Education), 3*(2), 174-182. doi:http://dx.doi.org/10.21831/reid.v3i2.17993

Introduction

Working, or having a career, is one of the major life tasks that affect the psychological wellness of an individual (Sweeney, 2009, p. 17). In relation to the development of psychological wellness studies, Myers, Sweeney, and Witmer (2000) mention that work is one of the factors that determine the well-being of one's mental health. They also imply that although statistically there is no direct correlation between them, the factor analysis result shows that career or having a job is fundamental to the acquisition of self-confidence in an individual. It is important that one can feel competent, esteemed, and belong to a positive identity in exploring his or her career options (Nauta, 2010, p. 12; Sweeney, 2009, p. 17).

Both career development and mental health play a role in helping an individual to achieve a more effective quality of life. Nauta (2010, p. 12), based on the Holland (1997) theory, states that an individual with a good understanding of his/her sense of self will

possess a crystalized vocational identity that allows him/her to explore his/her career options without difficulty. Such an individual will eventually be able to have a career that is sustained by competence, in addition to life satisfaction and an effective personality in his/her social life. This notion explains that a well-developed career relies on a good understanding of self-identity (in the context of realistic, investigative, artistic, social, enterprising, and conventional/RIASEC personality types), by which an individual will be able to determine his/her career direction. For this reason, it can be understood that the knowledge acquisition on career options in the early development stage is imperative to the improvement of quality of life during adulthood.

However, in reality, making a career decision can be a challenge. In addition to the limited jobs available in the market, competence, qualification, and interest also play an important role in what career a person may have. A survey conducted by the Organization for Economic Cooperation and Development (OECD & Asian Development Bank, 2015) shows that 6.2% of Indonesian population is unemployed. Although the number is slightly below the world's average unemployment (6.7%), it is relatively high when compared to other countries like Japan, the country with the lowest unemployment rate, where the number is twice less than Indonesia with 3.4%. Following Japan, the second, third, and fourth place are India, South Korea, and also China, respectively. These are the countries with rapid economic growth dominating the global market, and therefore, are able to create more jobs.

Career introduction is one of the factors affecting unemployment rates. On the other hand, job suitability and level of interest can also determine how long a person can last in his/her job. Oftenly, a person would quit a job in which they did well because the job does not fit his/her interest. In some instances, the lack of interest in the job could discourage workers to do their best. If this problem can be overcome, the number of people who quit or have no job can be reduced, resulting in a lower unemployment rate.

Funded by the Islamic Development Bank, the first, second, and third authors conducted a study in 2015 on the career development among lowergrade primary students as an effort to nurture career development among individuals from the early age. The first-year study was a preliminary research aimed at examining whether the lower-grade primary students in the *Daerah Istimewa* Yogyakarta (DIY), had acquired the adequate levels of knowledge and interest on various types of occupations. The six Holland's career categories comprising Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC) were used as a reference in measuring the levels of knowledge and interest among the subjects as they had been proven to possess a high reliability level in numerous settings (Nauta, 2010, pp. 17–18). The study found that the knowledge and interest levels of the students were in accordance with the theory of career interest by Holland, although they were varied from high to low. It was then concluded that the students needed additional support to explore and learn as many career options as possible. As a follow-up of the previous study, this second-year research focuses on the development of learning aid or media that aims to improve the lower-grade primary students' knowledge of various types of occupations.

One of the challenges in the previous study was how to make the students interested in the learning activities aimed at stimulating their knowledge acquisition of career options, especially as there were several jobs that they were not familiar with, which were understandably more difficult to comprehend due to the more complex nature of the jobs. Therefore, it is important that an appropriate and suitable media for lower-grade primary students be created to improve their career knowledge acquisition in a broader scale, at least within the proximal development zone.

Nowadays, the options for learning media development are endless, especially with the rapid development of technology. The opportunity to develop online-based media by using smartphone application is widely open and oftenly preferable, as it offers a quick and affordable access, allowing children to inde-

pendently use it at any time, in any place. Both computer and smartphone applications are fun and easy to use and produce. Nevertheless, it needs to be pointed out that advanced technology does not always have a good impact on children, especially when it comes to social interaction with their peers. Interaction is one of the most important elements in the cognitive, affective, and psychomotor development in children. According to a number of theorists, the interaction and relationship among children and their peers are just as crucial as those between children and their parents, in spite of the difference in quality and values (Piaget, 1965, in Sigelman and Rider, 2006, p. 408). While parents act as the authoritative figures in a child's life, children can learn to learn, respect each other, negotiate, and practice cooperative team-work with their peers as they have equal position and power. Piaget further emphasizes that peers have a distinct contribution to children development that parents cannot provide.

Based on the two points above, it must be acknowledged that there is a direct need for a learning media hardware that stimulates children's social interaction with their peers. One convenient example is the playing cards Quartets. Quartet is a game originated from the Netherlands used by children to learn words. However, in its development, the game has been greatly modified for numerous purposes. In the same vein, this study develops the Quartet playing cards to improve children's knowledge of career options. Thus, this study aims to (1) develop and (2) examine the feasibility of the Quartet playing cards as the media for improving the career knowledge acquisition of lower-grade primary students.

Career introduction and awareness in the early age are incredibly crucial for future career development. Knight (2015) states that introducing career options and information on tertiary education earlier in primary school can help students realize the important connection between education and vocational success for their future. This means that introducing career knowledge to primary students as early as they are at the lower grades would contribute to their success in the future.

The deliberate decision to use a game of playing cards as learning media was made to address the research subjects, i.e. early primary students who are still in the playing period. Among other educational toys for children such as puzzles, crossword puzzles, cards, pictures, movies and videos, and also interactive CDs, Quartet playing cards were chosen for several reasons.

The game is programmed to be played repeatedly and to engage children in a fun learning activity through the use of pictures to show various occupation types in the cards. This mechanism is designed in order to address the instrumental conditioning learning theory on how repeated activities can form a strong stimuli-response (S-R) connection, especially those involving emotion or happy feeling, where the stimuli tend to transform into a long-term memory (Santrock, 2008). This is supported by Garris, Ahlers, and Driskell (2002) who state that using toys or involving an element of fantasy in the learning activities allows students to feel more excited in the class.

Method

The activities in this second-year study involved developing a research product and conducting feasibility tests on the product. The process of developing the Quartet playing cards as the media of career exploration referred to ten-step research and development method proposed by Borg and Gall (1983), consisting of (1) research and information collection; (2) planning; (3) preliminary product form development; (4) preliminary field testing; (5) main product revision; (6) main field testing; (7) operational product revision; (8) operational field testing; (9) final product revision; and (10) dissemination and implementation. This second-year study adopted only the second to ninth steps, as the first step had been completed in the preceding study, while the final step is planned to be conducted in the third year.

The research instrument for data collection was a test resulted from the previous year's study (Ayriza, Setiawati, & Triyanto, 2016). The test was constructed according to the six Holland's Career Categories, i.e. realis-

tic, social, enterprising, and also conventional (RIASEC). There were a total of 60 items with each dimension containing 10 items.

Subsequently, the instrument underwent expert and construct validation. Two experts were involved in this study: an expert who has an extensive research experience in career development study; and an expert in primary school education. The expert on career development suggested that there were more varied alternative answer options for the test items, and to avoid similar answer alternatives/options for different statements. On the other hand, the expert on primary education found no significant problem except for the use of some technical terms. Once the test was revised based on the validators' advice, a construct validation was conducted by analyzing the six dimensions individually using the confirmatory factor analysis (CFA). The result shows that all test items in the dimensions refer to one factor. Each dimension was considered valid or significant if it had the Chi square probability of ($\chi^2 > 0.05$), as well as Root Mean Square Error of Approximation (RMSEA) or the average size of the expected difference per degree of freedom (df) in the population of less than 0.08.

The result of CFA is presented in Table 1. According to the table, each dimension of RIASEC corresponds to each latent variable and leans to one factor, meaning that it fulfills the requirement for construct validity. Meanwhile, instrument's realibility was tested using Alpha Cronbach formula and the result of reliability coefficient is 0.891.

Subjects

The population of this study was the lower-grade students of primary schools in *Daerah Istimewa* Yogyakarta (Special Regions of Yogyakarta), which consist of four regencies and one municipality. The cluster random sampling technique was used to establish the sample. Cluster referred to the regencies or municipality which have different characteristics, and random reffered to the technique used in selecting both the schools from the selected regencies and municipality, and the classes of the selected schools.

The research subjects were 266 primary school students of grade 1, 2, and 3. The numbers of students undergoing preliminary field testing, main field testing, and operational field testing were 12, 83, and 171 students, respectively. Details of the number and classification of the subjects are presented in Table 2.

Table 1. The result of the confirmatory factor analysis

Statistics	Dimensions					
	R	I	A	S	E	C
χ^2	34.03	35.30	16.04	41.49	34.97	39.44
df	26	26	15	29	27	32
significance (p)	0.13436	0.10535	0.26074	0.06235	0.13960	0.17136
RMSEA	0.023	0.025	0.019	0.027	0.023	0.020
Result	Fit	Fit	Fit	Fit	Fit	Fit

Table 2. The number of subjects in each field testing

Preliminary field testing	Primary school	SDN Samirono (Kota)					
	Grade	Grade 1		Grade 2		Grade 3	
	Total	4		4		4	
Main field testing	Primary school	SDN Karangmojo II (Gunungkidul)					
	Grade	Grade 1		Grade 2		Grade 3	
	Total	27		30		26	
Operational field testing	Primary school	SDN Kotagede (Kota)		SDN Sonosewu (Bantul)		SDIT Tunas Mulya (Gunungkidul)	
	Grade	Grade 1	Grade 2	Grade 2	Grade 3	Grade 1	Grade 3
	Total	29	27	30	23	30	32

Data Analysis Technique

The descriptive quantitative analysis technique was employed in this study. In addition, feedback and suggestions from the teachers and research assistants who were in charge of guiding and monitoring the field testing were compiled as a part of the qualitative data used as additional references to improve the research product.

Findings and Discussion

Findings

This section illustrates the findings of the current study which are based on the second to ninth step of Borg and Gall's ten-step R and D method. A brief summary of the finding of the first-year study is provided to give a comprehensive depiction of the study.

The first stage, research and information collection, shows that the career knowledge and interest of the subjects fit Holland's RIASEC construct theory. The second-year research is a follow-up of the previous study's findings.

The planning stage is manifested in developing the concept of Quartet playing cards based on Holland's RIASEC theory. At the top of the card, the word for the type of occupation is made bigger and bolder, while a description of the occupation is provided at the bottom. The cards are grouped into three volumes based on the difficulty levels (low, medium, and high), and four categories, i.e. task, tool, workplace/product/service, and working attire or attribute.

The stage of developing prelimary product form involves creating the product prototype as designed in the previous stage, and conducting feasibility test on it. Validity tests were performed by a professor with an extensive research on career development, and a Ph.D scholar on primary education, with the following results: (a) Several adjustments were made on the pictures, the size of the cards to fit the size of a child's palm, and on the thickness of the material to ensure the cards' durability; (b) adjusments were made on the measurement of the career knowledge instrument.

The next stage is the preliminary field testing. In this stage, four students of SDN Samirono (Yogyakarta) were randomly selected from each grade, making a total of 12 subjects. The result of the limited field testing is presented in Table 3.

Table 3 reveals that most of the first graders (75%) are able to play the game, find it easy to play, and will play it at home, while all second and third graders (100%) have no problem and respond positively to all of the test items. All subjects from the three grades (100%) agree that the playing card helps them learn many types of occupationss and enjoys playing the game. However, about 25% of the first graders still have difficulty in playing the game, find it hard to play, and will not consider playing it at home. In addition, the minority also does not find it easy to learn the characteristics of the occupations. Unlike the second and third graders, the first graders show more enthusiam for obtaining the goal of the Quartet Career Cards as learning media. Additionally, during the limited field testing, the research team received feedback and suggestions from the teachers and research assistants, including to make the font size bigger, the colors brighter, as well as adjustment on the pictures to reflect the geographical condition of where the students live.

Table 3. The result of limited field testing

Primary school	**SDN Samirono**			**Mean**
Grade	Grade 1	Grade 2	Grade 3	
Subjects	4 students	4 students	4 students	
Able to play the game	75%	100%	100%	92%
Find the game easy to play	75%	100%	100%	92%
Will play at home	75%	100%	100%	92%
Learn many types of occupationss	100%	100%	100%	92%
Learn the characteristics of the occupations	75%	100%	100%	92%
Enjoy the game	100%	75%	100%	92%

In the next stage, the research team conducted revision to the main product based on the recommendation and feedback from the students and teachers in stage four. The result of the revision was the final draft of main product that was ready for the main field testing.

The main field testing was conducted in the same mechanism as the preliminary field testing in stage four, but with larger subjects. There were 27 grade 1 students, 30 grade 2 students, and 26 grade 3 students of SDN Karangmojo Gunungkidul. The data are presented in Table 4.

As illustrated in Table 4, the results of extended main field testing are varied among students of grade 1, 2, and 3. Nearly all of the students (90%) are able to play the Quartet Career Cards game and find it easy to play. About the same number (94%) of students will play the game at home, and manage to learn the types (92%) and also characteristics (94%) of the occupations from playing it. On average, there are 99% of students who enjoy playing the game. Overall, 93% of the students achieve the goal of playing the game, as targeted by the research team.

In the next stage, operational product revision was made according to the feedback and suggestions of the subjects during the extended main field testing, particularly on the colors in the type of occupation and the answer choice. Moreover, the colors in the low, medium, and high levels were changed into red, blue, and green, respectively. Once the revision was made, it was concluded that the developed learning media was ready for operational field testing.

The operational field testing was done in one primary school in the municipality of Yogyakarta, and two other schools in the *DIY* regencies. They were SD Kotagede (Municipality of Yogyakarta) with 29 first graders and 27 second graders; SD Sonosewu (Bantul Regency) with 30 second graders and 23 third graders; and SDIT Tunas Mulia (Gunungkidul Regency) with 30 first graders and 32 third graders. The result is presented in Table 5.

Table 5 shows that the data obtained from operational field testing are varied. In general, the first grade students of the three schools are able to play the game (90%), find the game easy to play (91.5%), will play the game at home (91.5%), learn many types (83%) and characteristics (76.5%) of the occupations, and enjoy the game (95%). On the other hand, 85% second graders are able to play the game and find the game easy to play,

Table 4. The result of extended main field testing

Primary school	**SDN Karangmojo II**			**Mean**
Grade	Grade 1	Grade 2	Grade 3	
Subjects	27 students	30 students	26 students	
Able to play the game	89%	90%	92%	90%
Find the game easy to play	89%	90%	92%	90%
Will play at home	100%	87%	96%	94%
Learn many types of occupationss	100%	77%	100%	92%
Learn the characteristics of the occupations	100%	87%	96%	94%
Enjoy the game	100%	97%	100%	99%

Table 5. The result of operational field testing

Primary school	**SDN Kotagede 1**		**SDN Sonosewu**		**SDIT Tunas Mulia**		**Mean**
Subjects	29	27	30	23	30	32	
Grade	1	2	2	3	1	3	
Able to play the game	83%	89%	81%	100%	97%	91%	90%
Find the game easy to play	86%	89%	81%	100%	97%	91%	91%
Will play at home	90%	100%	97%	100%	93%	97%	96%
Learn many types of occupations	86%	96%	100%	100%	80%	97%	93%
Learn the characteristics of the occupations	83%	93%	97%	100%	70%	100%	91%
Enjoy the game	90%	93%	100%	100%	100%	100%	97%

99% will play the game at home, 98% learn many types of occupations, 95% learn the characteristics, and 97% enjoy the game. Finally, among grade 3 students, 95.5% students are able to play the game and find it easy to play, 98.5% will play it at home and learn about many types of occupations, while 100% students both learn the characteristics and enjoy the game.

The operational field testing reveals that grade 3 students have the highest achievement of the research goal (98%), followed by the second graders (93%), and the first graders (88%). On average, there are 93% lower-grade primary school students who achieve the goal of Quartet Card Careers as targeted by the research team.

The ninth and final stage in this research was final product revision. This stage resulted in a suitable final product of career exploration for children aimed at improving their career knowledge.

Discussion

This study is a follow up of the first-year study on the exploration of career interest and knowledge construct using quantitative analysis. The first-year study shows that both the career interest and knowledge of the lower-grade primary students in *Daerah Istimewa* Yogyakarta are society-oriented, and that they correspond well to Holland's theory of six career categories (RIASEC). The fact that students' career knowledge is society-oriented implies that their career interests are limited to the social scope, as well.

As a result, children's career development may be disrupted, especially when there is no intervention. Therefore, the second-year study was aimed at improving children's career knowledge based on Holland's RIASEC theory through Quartet Career Cards as the learning media specifically developed for that purpose.

The decision to conduct intervention on the lack of career knowledge among lower-grade primary students is based on the research by Xu, Hou, and Tracey (2014, p. 654) in China, which revealed that the lack of self-exploration and environment exploration was caused by the lack of information or knowledge of exploring career options, as well as the lack of efforts or supporting facilities in the career exploration process. In that case, intervention is imperative to make improvements on the children's career knowledge.

The final product of the developed learning media in this study is modified Quartet cards containing pictures and information aimed at lower-grade primary students' career development. It is expected that students can acquire wider knowledge and develop their career interests, so that they are well-informed and ready to make a decision on their vocational preference when they grow up. In addition, the game is also intended to increase the players' interpersonal relationships and give them pleasure and enjoyment when playing it.

The notion is in line with the study of Garris et al. (2002) who state that games can be used as a part of learning activities to make students enjoy the learning process more due to the element of fantasy. Moreover, Parker and Lepper (1992) find that learning in an environment which involves an element of fantasy is more beneficial to students than one conducted in other conditions.

All field testing, whether preliminary, main, or operational, shows that more than 90% students are able to play Quartet Career Cards, find it easy to play, enjoy the game, and will play it again at home. The learning aspect of the game is aimed at helping students improve their career knowledge. This is evident in how students manage to learn more types of occupations, what the jobs entail, and the relevant tools, setting, attire or attributes required for particular jobs. For instance, they children learn that *caping* (a traditional wide cone-shaped hat made from bamboo) is an attribute associated with Indonesian farmers as they do not have special attires for their job.

On the other hand, the discussion on career exploration as early as primary school was in accordance with the study conducted by Magnuson and Starr (2000). They argue that making simple decisions during early childhood such as choosing what food to eat, toys to play, clothes to wear, or things to do in their daily lives will help the children have personal preferences. In the future, the preferences will manifest in the formation of in-

dividual autonomy that helps them to make decisions and life choices, including in determining what career they would have as an adult.

Based on Piaget's cognitive development theory, the lower-grade primary students are in the concrete operational stage, which is marked by how they are not dominated by perception and rely on experience to guide them, in addition to the extraordinary cognitive development and formative stage in the formal education setting (Schunk, 2012). The concrete operational stage includes the ability to classify, combine, and compare. In this stage, children are also able to understand the connection and to make sense of a series of events (Hill, 2012). The theory implies that as educators, teachers should be able to provide the appropriate learning style and environment according to the students' cognitive development so that students are encouraged to explore and actively participate in social interaction. In relation to learning environment, teachers are also responsible for giving new stimuli for students' cognitive construct to stimulate their development through assimilation and accommodation.

As learning media, the Quartet Career Cards allow the lower-grade primary students to explore a variety of possible career options for their future, as well as to engage in an active participation by interacting with their peers. As a result, the game acts as a stimulus for the environment that simultaneously constructs the children's cognitive structure with career knowledge. Taveira, Silva, Rodriguez, and Maia (1998, p. 90) emphasize the significance of early career exploration for children at the primary school age to support and foster the children's proper development.

Based on the research development, field testing, findings, discussion, relevant theories and previous studies, it can be concluded that the Quartet Career Card game is a contributing factor in children's career exploration process, as it can help improve the career knowledge of lower-grade primary students. This implies that Quartet Career Card can be recommended as media in career guidance activities to expand and enhance children's career knowledge.

Conclusion

The developing of the Quartet Career Cards game is aimed at supporting children's future career development by improving their knowledge on possible career options. While the product development process is based on Borg and Gall's (1983) ten-step model, the concept of the cards itself relies on Holland's theory of the six Career Categories, involving realistic, investigative, artistic, social, enterprising, and conventional.

The Quartet Career Cards consist of a picture and information on the types of occupations and each of their tasks, tools, workplace, products or services, as well as the attributes and work attires. There are three levels of difficulty ranging from high, medium, to low. The cards are proven suitable and feasible to be used by lower-grade primary students, based on a series of field testing, as well as validity tests conducted by theory and media experts.

Acknowledgment

The authors express gratitude to the Islamic Development Bank and Universitas Negeri Yogyakarta for the financial support in conducting this research, under the grant of *Penelitian Unggulan Perguruan Tinggi Tahun Anggaran* 2017 No. 19/Penel./ P.UPT/UN34.21/ 2017.

References

Ayriza, Y., Setiawati, F. A., & Triyanto, A. (2016). Career interest and knowledge of lower grade students of primary school. In *International Conference of Computer, Environment, Social Science, Engineering and Technology (ICEST 2016)* (p. May 23-25th). Medan, Indonesia.

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

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441–467. https://doi.org/10.1177/1046878102238607

Hill, W. F. (2012). *Theories of learning: Teori-teori pembelajaran, konsepsi, komparasi dan signifikansi* (5th ed.). (M. Khozim, Trans.). Bandung: Nusa Media.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources. https://doi.org/10.1016/0022-4405(74)90056-9

Knight, J. L. (2015). Preparing elementary school counselors to promote career development: Recommendations for school counselor education programs. *Journal of Career Development*, *42*(2), 75–85. https://doi.org/10.1177/0894845314533745

Magnuson, C. S., & Starr, M. F. (2000). How early is too early to begin life career planning? The importance of the elementary school years. *Journal of Career Development*, *27*(2), 89–101. https://doi.org/10.1177/089484530002700203

Myers, J. E., Sweeney, T. J., & Witmer, M. (2000). The wheel of wellness counseling for wellness: A holistic model for treatment planning. *Journal of Counseling & Development*, *78*(3), 251–266. https://doi.org/10.1002/j.1556-6676.2000.tb01906.x

Nauta, M. M. (2010). The development, evolution, and status of Holland's theory of vocational personalities: Reflections and future directions for counseling psychology. *Journal of Counseling Psychology*, *57*(1), 11–22. https://doi.org/10.1037/a0018213

OECD, & Asian Development Bank. (2015). *Education in Indonesia: Rising to the challenge*. Paris: OECD. https://doi.org/10.1525/as.1951.20.15.01p0699q

Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, *62*(4), 625–633. https://doi.org/10.1037/0022-3514.62.4.625

Santrock, J. W. (2008). *Educational psychology* (3rd ed.). New York, NY: McGraw-Hill.

Schunk, D. H. (2012). *Learning theories: An educational perspective*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

Sigelman, C. K., & Rider, E. A. (2006). *Life-span human development* (5th ed.). Belmont, CA: Thomson Wadsworth.

Sweeney, T. J. (2009). *Adlerian counseling and psychotherapy: A practitioner's approach* (5th ed.). New York, NY: Taylor & Francis. https://doi.org/10.4324/9780203886144

Taveira, M. D. C., Silva, M. C., Rodriguez, M. L., & Maia, J. (1998). Individual characteristics and career exploration in adolescence. *British Journal of Guidance & Counselling*, *26*(1), 89–104. https://doi.org/10.1080/03069889808253841

Xu, H., Hou, Z.-J., & Tracey, T. J. G. (2014). Relation of environmental and self-career exploration with career decision-making difficulties in Chinese students. *Journal of Career Assessment*, *22*(4), 654–665. https://doi.org/10.1177/1069072713515628

SUBJECT INDEXES

AUTHOR INDEXES

AUTHORS' BIOGRAPHY

Agus Triyanto. Currently works as a lecturer in the Department of Guidance and Counseling, Faculty of Education, Universitas Negeri Yogyakarta. He attained bachelor of guidance and counseling in 2003 from Universitas Negeri Yogyakarta. His master degree was attained in 2011 from Universitas Negeri Malang at the same major.

Alita Arifiana Anisa. She was born on April 19, 1991. She attained her bachelor degree in Accounting Education in 2013 from Universitas Negeri Yogyakarta. She continued her study at the same university and attained her master in Educational Research and Evaluation in 2015.

Amipa Tri Yanti Nadapdap. Was born in Padangsidempuan, North Sumatra, Indonesia on April 5, 1992, she attained her bachelor degree in physics education from Universitas HKBP Nommensen, Medan, North Sumatra, in 2014. Her master degree on the same major was attained from Universiats Negeri Yogyakarta in 2017.

Claver Nzobonimpa. Claver Nzobonimpa completed his master degree in Social Science Education at Universitas Negeri Yogyakarta, Indonesia, in 2016. He currently works as a lecturer of Sociology in the Department of English Language and Literature, Faculty of Languages and Social Sciences, Burundi National University.

Dian Normalitasari Purnama. Was born on November 19, 1990 in Gunungkidul, Yogyakarta, she currently pursues her doctoral degree in Universitas Negeri Yogyakarta, majoring Educational Research and Evaluation. Her master and bachelor degrees were attained in the same university, majoring Educational Research and Evaluation and Accounting Education, respectively.

Djemari Mardapi. Was born on January 1, 1947, he is a professor at Universitas Negeri Yogyakarta, Indonesia. He currently works as a lecturer at the Faculty of Engineering and the Graduate School of Universitas Negeri Yogyakarta. He obtained a bachelor degree in Electrical Engineering Education in 1973, from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta). In 1984, he achieved his master degree on Educational Research and Evaluation from Universitas Negeri Yogyakarta. He is an alumnus of IOWA University in 1988, majoring in Educational, Measurement, and Statistics and obtained a Ph.D.

Edi Istiyono. Was born on March 7, 1968, he currently works as a lecturer in physics education department and educational research and evaluation study program at Universitas Negeri Yogyakarta. Having research expertise in physics education research and evaluation, he attained his bachelor degree in physics education in 1992, from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta). In 1999, he graduated from Universitas Gadjah Mada attaining his master degree on physics. In 2014, he attained his doctoral degree in educational research and evaluation from Universitas Negeri Yogyakarta.

Farida Agus Setiawati. Currently works as a lecturer in the Department of Psychology, Faculty of Education, Universitas Negeri Yogyakarta. Her bachelor degre in psychology was attained in

1996 from Universitas Gadjah Mada. Her master degree in psychometry was attained in 2004 from the same university. She attained her doctoral degree in Educational Research and Evaluation from Universitas Negeri Yogyakarta in 2013.

Kartianom. Was born in Bau-Bau, Buton Island, Indonesia, on November 17, 1990, he is pursuing his doctoral degree in Educational Research and Evaluation in Universitas Negeri Yogyakarta. He completed his master education in the same major in the same university in 2017. Meanwhile, his bachelor degree in Mathematics Education was attained in 2014 from Universitas Datanu Ikhsanuddin, Bau-Bau.

Kumaidi. Currently works as a lecturer of psychometrics and psychological statistics in the Faculty of Psychology of Universitas Muhammadiyah Surakarta. He attained his bachelor degree in 1976 from Universitas Negeri Yogyakarta, Indonesia, majoring Mechanical Engineering Education. His master degree was attained in 1984 from University of Iowa, USA, majoring Educational Measurement and Statistics, and his doctoral degree was attained in 1987 in the same major and university.

Moh Khoerul Anwar. Starts working as a lecturer in the department of Islamic Guidance and Counseling in Universitas Islam Negeri Sunan Kalijaga Yogyakarta since the beginning of 2017, he attained his bachelor and master degrees in Universitas Negeri Yogyakarta in the same major: Guidance and Counseling, in 2014 and 2016, respectively. His experience in becoming a former teacher in Bina Anggita, a school for children with special needs, and a former mentor in *Madrasah Diniyah Takmiliyah* Al Ikhlash Samirono, a non-formal Islamic education program for children in primary school ages, makes him own special competence in educating children.

†Muhammad Mustaghfirin Amin. He worked as the Director of Technical and Vocational Education, at the Directorate General of Secondary Education, the Ministry of Education and Culture since 2013 until he passed away in September 2017. He pursued his bachelor degree in Electrical Engineering in Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta), his master degree in Management in Universitas Gajayana Malang, and his doctoral degree in Educational Research and Evaluation in Universitas Negeri Yogyakarta.

Nanang Erma Gunawan. He currently works as a lecturer in the department of Educational Psychology and Guidance in Universitas Negeri Yogyakarta. He attained his bachelor degree in Guidance and Counseling from Universitas Negeri Yogyakarta in 2008. Meanwhile, he has attained his master degree in 2013 in Ohio University, USA, majoring Clinical Mental Health and Rehabilitation Counseling.

Nugraheni Dwi Budiarti. Was born in Ngawi, East Java, on December 10, 1976, she currently works as a teacher at Sekolah Dasar Muhammadiyah Banguntapan. Besides, she completed her master degree in Primay Education at Universitas negeri Yogyakarta in 2018.

Oscar Ndayizeye. Oscar Ndayizeye was born on February 17, 1983 in Kayogoro/Makamba Province, in southern Burundi. In 2012, he graduated as an Agrégé of TEFL at the Institute of Applied Pedagogy, University of Burundi. In 2015, he was admitted in Educational Research and

Evaluation study program at Universitas Negeri Yogyakarta, Indonesia. He completed his study in June 2017 and attained a Master Degree in Education/Evaluation concentration.

Soenarto. Was born on 4 August 1948, he works as a senior lecturer in the Faculty of Engineering and the Graduate School of Universitas Negeri Yogyakarta, Indonesia. He attained his bachelor degree in 1974 from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta) majoring in Electrical Engineering Education. In 1984, he attained his Master of Science degree in Industrial and Vocational Education, from State University of New York, USA. In 1987, he also achieved a Master of Arts degree in Educational Program Evaluation from Ohio State University, USA. His Doctor of Philosophy degree in Industrial Vocational Education was attained in 1988 from Ohio State University, USA.

Yulia Ayriza. Was born on July 3, 1959, she currently works as a lecturer in the department of Psychology, in the Faculty of Education and the Graduate School of Universitas Negeri Yogyakarta. Her bachelor and master degrees in Psychology were attained from Universitas Gadjah Mada, Indonesia, in 1983 and 1995, respectively. Besides, she also attained a doctoral degree from the School of Social Science, Universiti Sains Malaysia, in 2013.

Zamroni. Currently works as a senior lecturer in Universitas Negeri Yogyakarta in the expertise field of educational sociology. He completed his bachelor degree in Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta) focusing on economic enterprise, and his master and doctoral degrees in Florida State University, United States of America achieving M.A. and Ph.D. He has been being active in publication since he started working in Universitas Negeri Yogyakarta in 1974.

SUBMISSION GUIDELINES

- The manuscript submitted is a result of an empirical research or scientific assessment of an actual issue in the area of 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 including tables, figures, references, and captions. Manuscripts that greatly exceed this will be critically reviewed with respect to length. (A4; margins: top 3, left 3, right 2, bottom 2; double columns [Except in Abstract: single column]; single-spaced; font: Garamond, 12).
- Manuscripts should be compiled in the following order: (1) title; (2) abstract; (3) keywords; (4) main text: introduction, method, findings and discussion, conclusion and 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].
- Authors' biography is written in the form of narration, including author's full name, place and date of birth, educational qualification/information started from bachelor degree (S1) until the latest educational degree, the affiliation in which the author is currently working, phone number, and email address.
- All Author(s)' names and identity(es) must be completely embedded in the form filled in by the corresponding author: email; affiliation; and each author's short biography (in the column of 'Bio Statement'). [if the manuscript is written by two or more authors, please click 'Add Author' in the 3rd step of 'ENTER METADATA' in the submission process and then enter each author's data.]
- All correspondences, information, and decisions for the submitted manuscripts are conducted through the email/s used for the submission.
- Word template is available for this journal. Please visit the journal's homepage at https://journal.uny.ac.id/index.php/reid
- If you have submission queries, please contact *reid.ppsuny@uny.ac.id* or *reid.ppsuny@gmail.com*