



Validity and Reliability of Development of Self-confidence Assessment Instruments for Students on Chemistry Subject

Noviashri Pramesty^{1✉}, Ani Rusilowati², Muhammad Khumaedi³

¹SDN 02 Rembul Kabupaten Pemalang

^{2,3}Universitas Negeri Semarang, Indonesia

Article Info

History Articles

Received:

23 December 2021

Accepted:

22 January 2022

Published:

30 March 2022

Keywords:

Instrument,
assessment, inclusive,
equalization

Abstract

This research aimed to develop a cognitive assessment instrument for learners with special needs in inclusive elementary schools, focused on slow learners. The type of the research used is R&D (Research and Development) with the development model proposed by Borg & Gall. Sukmadinata modified the design from Borg & Gall in three main steps, that is (1) preliminary study; (2) product development; and (3) product testing. The subjects of this study were 3rd Grade students of SDN 14 Mulyoharjo. Research data obtained through interviews, observations and tests. The instrument developed was made in two question packages. Based on the result of the study, it was found that the assessment instrument aimed at inclusive elementary students was declared feasible. This is indicated by content validity validity, which was 0.80 to 1.00, the reliability between experts that resulted as much as 0.796 is categorized as good, and the reliability of large-scale test instruments belong to the very reliable category as indicated by the Kuder Richardson (KR) 20 value of 0.983 and 0.976. The equivalence test was carried out using the mean equalizing method and yielded a value of 8.71. Therefore, it can be concluded that the assessment instrument intended for students in inclusive elementary schools has met the specified criteria values so that it is feasible to use. The implication of the assessment instrument is that it can facilitate and assist teachers in inclusive primary schools in carrying out assessments.

✉Correspondence Address :

SDN 02 Rembul, Jalan Randudongkal-Kreyo Km. 4,

Pemalang Indonesia 52353

E-mail : noviashripramesty@gmail.com

p-ISSN 2252-6420

e-ISSN 2503-1732

INTRODUCTION

Special education is part of the Indonesian National Education System which is specifically described in Act no 20 of 2003 Article 32 Paragraph 1. In providing educational services to learners with special needs, the treatment cannot be generalized, because each type of problems encountered has its own characteristics. So it must involve various special skills according to the needs of each type of abnormality in the learners with special needs. Not all learners with special needs require services in special schools. Special schools are designed for learners with special needs who have severe disabilities or other reasons that made them unable to study together with other learners (Malak, 2013). However, there are also learners with special needs who still can study together with other normal learners but only need special services.

The Inclusive School is one of the educational programs designed to provide services for learners with special needs so that they can attend school with other normal learners through formal channels. According to Abdurrahman (2010: 27), inclusive education was born as a form of dissatisfaction with the implementation of education for learners with special needs using a segregation system. The segregation system is a school administration system intended for learners with disabilities or learners with special needs.

Basically, the curriculum model in implementing inclusive education is categorized into 3, namely the general curriculum, modified curriculum, and individualized curriculum (Garnida, 2015: 107). The general curriculum is the curriculum used in regular schools in general. While the modified curriculum is a regular curriculum that is modified and adapted to the abilities and characteristics of learners with special needs (Schuelka, 2018). Then the individualized curriculum is a curriculum aimed at learners with special needs who cannot take lessons in the general curriculum or modified curriculum (Salomi, 2018). The curriculum used in the

implementation of inclusive education is currently mostly using the regular curriculum that applies as in public schools. However, because the variety of obstacles experienced by students with special needs varies greatly, ranging from mild, moderate, to severe, in its implementation, the regular curriculum needs to be modified in such a way that it fits the needs of students. Shaw, et al., (2005), explained that basically there are no so-called normal children, but there are children with individual differences.

Based on the research result carried out by Haryono, Syaifudin, and Widiastuti (2015: 124-125), schools providing inclusive education in Indonesia, especially in Central Java have not implemented an inclusive education system as it should. The results of the study show that in Central Java Province there are 519 inclusive schools and it is known that (1) student management in inclusive education in Central Java Province is considered to be still lacking; (2) curriculum management in inclusive education in Central Java Province is considered lacking; (3) educational personnel management in inclusive education in Central Java Province is considered lacking; (4) facilities and infrastructure management in inclusive education in Central Java Province is considered lacking; (5) financing management for inclusive education in Central Java Province is considered lacking; and (6) management in environmental and special services in inclusive education in Central Java Province are considered lacking.

In essence, in inclusive schools, in addition to using a curriculum that is adapted to the students' condition, there are also special supervisory teachers who assist teachers in the classroom learning process. Usually special supervisory teachers come from special schools which have the ability to handle students with special needs. They have their own learning methods that are applied to students with special needs according to the specificity of each individual. However, not all inclusive schools have special supervisory teacher, so classroom teachers must put extra effort in

providing learning services to students in these inclusive schools. As happened in the elementary schools of SD Negeri 05 Mulyoharjo and SD Negeri 14 Mulyoharjo, Pemalang Regency. This happened because the school was unable to finance the procurement of special supervisory teacher.

The absence of modifications to the curriculum and learning in schools providing inclusive education also has an impact on the learning outcome assessment system for learners with special needs (Chineye, 2012). Assessment for students with special needs cannot be generalized with normal students (Miller, et al., 2015). The ability of students with special needs is clearly different from normal students. Moreover, in inclusive schools where most of the students with special needs are slow learners, their learning abilities, especially in the cognitive aspect, are considered lagging behind or below normal students, thus requiring differences in the learning outcome assessment (Sangeeta, 2009).

Haug (2017: 206-217) explained that one of the main challenge in developing inclusive education, is in terms of assessment, inclusive education must pay attention to all students with special educational needs and not only students with disabilities, which is the perspective that dominates today. Assessment is one of the important factors in learning to measure students' development and achievement (Alfianto, et al., 2015). In addition, according to (Arifin, 2014) Assessment is a systematic and continuous process or activity to collect information about the process and learning outcomes of students in order to make decisions based on certain criteria and considerations. Therefore, to obtain valid learning outcomes, it is necessary to have an assessment that is carried out correctly, systematically, and in accordance with the needs of students.

Based on the research result that has been carried out by Pramesty (2017), it showed that learning in inclusive schools at SD Negeri Slerok 2 Tegal City used the same curriculum and assessment as regular schools in general.

Even though the assessment for students with special needs cannot be generalized with normal students. In addition, the researcher has made observations at several inclusive schools in Pemalang Regency, namely SD N 14 Mulyoharjo and SD N 05 Mulyoharjo. The conditions are the same, learning in both schools uses the curriculum and methods like regular schools in general. There is no modification either in the learning process, or in the assessment system. Therefore, the development of assessment instruments in inclusive schools needs to be carried out to facilitate classroom teachers in providing assessments of learning outcomes for learners with special needs.

Based on the description above, the researchers are interested in developing cognitive assessment instruments for students with special needs in inclusive elementary schools that are focused on Slow Learners. This is because most of the students with special needs who study in inclusive schools are learners with special needs types who have cognitive learning difficulties, while learners with special needs in the more severe types are directed to study in Special Schools.

METHOD

The research method used for this research is R&D (Research and Development). This study used the development model proposed by Borg & Gall. This study developed a cognitive assessment instrument for slow learners with special needs in inclusive schools. Sukmadinata modified and presented an outline of the design by Borg & Gall. Sukmadinata (2012: 184) conveyed the research and development model in three main steps, namely (1) preliminary study; (2) product development; and (3) product testing. This model does not eliminate the existing steps. This model has also been tested by being used in several national-level development studies.

The test in this study was carried out in three stages, namely expert test, field test, and also individual test. This is in accordance with

Zuliani's research (2017: 49), which stated that product development trials are carried out through 3 stages, namely expert test, individual test and field test. The test is intended to obtain a suitable instrument for use in research (Kass, & Tinsley, 2009).

The subject of this study were 3rd grade elementary school students of SD N 14 Mulyoharjo. The data collection techniques in this study include observation, interview, and test. The instruments used in this study were (a) interview guides; (b) expert validation sheet; (c) cognitive assessment instrument in Inclusive Schools. The cognitive instrument used test instruments which are used to assess students in inclusive elementary schools (Qomari, 2008).

The data analysis used in this research activity is qualitative data analysis and quantitative descriptive data analysis (Yusuf, 2017). Qualitative data analysis was used to answer how to develop questions for students with special needs in inclusive schools. Quantitative descriptive data analysis was used to analyze the data obtained in the form of

expert validation assessment scores and field tests to measure the level of validity and reliability of the instrument, as well as test equating using the mean method classical test theory.

RESULT AND DISCUSSIONS

In this study, the developed instrument modifies the part of student learning outcomes evaluation. The modifications made are including grids and questions. The modified components are the question indicators and the measured cognitive level. The form of the question is multiple choice. This study uses 4 answer choices which are tests given at the elementary school to high school level. The answer choices consist of one correct answer with a score of 1 and the wrong answer with a score of 0. The material raised is Theme 3, Sub-theme 1 in 3rd grade which contains 4 subjects namely Civics, Bahasa Indonesian, Mathematics, and Cultural Arts and Crafts. The modifications made is presented in the following Table 1 and Table 2:

Table 1. Identitiy Component

Component	Content	Description
School Type/Level	Elementary School...	Educational institution name
Major	3rd grade	Class grade/level
Subject	Theme 3 Sub Theme 1	Compatible theme
Academic Year	2021/2022	According to the current academic year
Curriculum	Kurikulum 2013	Effective curriculum
Time Allotment	35 minutes	One unit hour for low class
Number of Questions	15 questions	Questions for learners with special needs and normal students are made differently but with the same amount/number of questions
Form of Questions	Multiple Choice	The form of the questions is made the same, but for the questions for students with special needs, the size of the letters is enlarged and lots of pictures are added as explanations

Table 2. Matrix Components

Component	Content	Description
Basic Competence	The basic competence is according to the syllabus	Same with the regular curriculum
Materials	The materials are in accordance to the syllabus and lesson plan	Same with the regular curriculum
Question Indicator	There are 2 packages namely Package A dan Package B	Package A is for students with special needs and Package B is for normal students. The question indicators are adjusted to the measured cognitive level.
Cognitive Level	Package A including the cognitive levels of C1, C2, C3, C4 and package B including the cognitive levels of C2, C3, C4, C5, C6. C2 and C3 in package B, and C3 and C4 in package A are the anchor points or shared points.	C1 and C2 are lower-level cognitive level or also known as Lower Order Thinking Skills (LOTS). C3 is the intermediate cognitive level or also known as Middle Order Thinking Skills (MOTS). While C4, C5, and C6 are high-level cognitive levels or also called Higher Order Thinking Skills (HOTS).
Question Number	Number 1 through 15	The question number in both packages are in the same order.

The questions are made in two packages with the same number of questions, that is 15 items, including in Package A and Package B. Package A is used to measure the learning outcomes of learners with special needs of the Slow Learner type, while Package B is used to measure the abilities of normal students. The difference between the two test packages lies in the question indicators and cognitive level in each question item.

Validity and Reliability of Cognitive Assessment Instrument for Slow Learners with Special Needs in Inclusive Elementary Schools

The content validity test by experts was used to see the completeness of the instrument, the suitability of the content, and the questions construction (Ogasawara, 2001). Quantitative summary data that presents the coefficient of expert agreement are presented in the following Table 3:

Table 3. Experts' Agreement Coefficient

Question no	Aiken's V Index	Criteria	Conclusion
2. 3. 7. 8. 9	1.0	Valid	Feasible
4. 6	0.875	Invalid	Minor Revision
1. 5. 10	0.8125	Invalid	Minor Revision

Based on the content validity analysis result using the Aiken's V formula, an information was obtained that 5 items out of 10 components or indicators assessed obtained a V index of 1.0 and belongs to valid category

(Akbar, 2013). While 2 items with a V index of 0.875 and 3 items with a V index of 0.8125 are not valid, because the minimum value of index V according to Aiken's V table is 0.88, in accordance with what is stated in an article by

Pratiwiningtyas (2017: 7), that the prerequisite items having an index (v) 0.300 is considered valid and can be used for the next stage, while the prerequisite items that are less valid are corrected so that there are slight revisions to these items.

The revisions include (1) item indicator number 1, namely regarding the question grid; (2) indicator item number 4 regarding scoring guidelines; (3) item indicator number 5 regarding the question items; (4) indicator number 6 regarding the suitability of the question with the answer key; and (5) indicator item number 10 regarding the function of distractors in answers.

Suggestions for improvement from experts are used as the basis for improving the assessment instrument at the one-to-one test stage. The one to one test stage receives input from users (learners) in terms of the

components of the questions they are answering. The researcher asked 3 learners with special needs from Slow Learners type and 3 normal students from low, medium, and high ability levels as testees. According to input from students with special needs, the Slow Learner type, the letter size in the Package A questions was changed from 12 to 14 so that they were more clearly readable. Meanwhile, from normal students, there was no input regarding the questions they had been working on.

The reliability tested is inter-rater reliability (among experts) and field trial/large scale reliability. The reliability test was analyzed using SPSS version 23. The results of the agreement reliability test from 4 experts after being analyzed by Intraclass Correlation Coefficients (ICC) were as follows:

Table 4. ICC Reliability Test Result

	Intraclass Correlationb	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.493a	.177	.805	4.892	9	27	.001
Average Measures	.796c	.462	.943	4.892	9	27	.001

Based on Table 4, the average agreement between raters can be seen on the average measures, which is 0.796. According to Mehta, et al. (2018: 2735) ICC is used to determine the consistency of a data, where a higher ICC indicates better consistency. Based on the results of the ICC value obtained, it can be concluded that the agreement between raters is at a high or good level, because it is in the range between 0.76 – 0.90 (Ghozali, 2016).

In addition, the large-scale instrument reliability test analyzed using Kuder Richardson (KR) 20 obtained the results as shown in the following Table 5:

Table 5. Package A Questions Reliability Reliability Statistics

KR20	Category
0.983	Very Reliable

Based on the output in the Reliability Statistics table with a Kuder Richardson (KR) 20 value of 0.983, it shows that the instrument used is very reliable because it is greater than 0.6. Furthermore, the reliability test on Package B questions can be seen in the following Table 6:

Table 6. Package B Questions Reliability Reliability Statistics

KR20	Category
0.976	Very Reliable

Based on the output in Table 6, it can be seen that with a Kuder Richardson (KR) 20 of 0.976, the instrument used is declared to be very reliable. This reliability criterion is based on the criteria mentioned in a research article by Soleh (2017: 75), stating that the value 0.60 – 1.00 is included in the very reliable category. In addition, Wells & Wollack (2003) revealed

that the reliability value has exceeded 0.7 so that the instrument can be used for assessment sheets in classroom learning.

Characteristics of Cognitive Assessment Instruments for Slow Learners with Special Needs in Inclusive Elementary Schools

The characteristics of the developed assessment instrument can be identified by measuring the difficulty level of the items and the discriminating power of the questions. Good test questions must be valid and reliable. In addition, the test questions must have a good level of difficulty and discrimination. The level of difficulty and discrimination are characteristics of test items (Kim & Cohen, 2002). Measurement of the level of difficulty is used to determine the degree of difficulty of the items. While the power of difference is used to determine the extent to which the items are able to distinguish students who have high abilities from students who have low abilities. The discriminating power and difficulty level were analyzed using Microsoft Excel. The results of the discriminating power test and the difficulty level of the two test packages is presented in the following data:

Table 7. Difficulty Level Test of Question Package A

Question Number	Difficulty Level	Category
1	0.71	Easy
2	0.57	Medium
3	0.43	Medium
4	0.29	Hard
5	0.57	Medium
6	0.57	Medium
7	0.29	Hard
8	0.71	Easy
9	0.71	Easy
10	0.43	Medium
11	0.43	Medium
12	0.29	Hard
13	0.71	Easy

Question Number	Difficulty Level	Category
14	0.57	Medium
15	0.57	Medium

Based on the difficulty level test table on Package A questions, the results obtained are 4 items belong to the easy category, 8 items in the medium category, and 3 items in the difficult category. Then for the level of difficulty in Package B questions, it can be seen in the following Table 8:

Table 8. Difficulty Level Test of Question Package B

Question Number	Difficulty Level	Category
1	0.71	Mudah
2	0.43	Sedang
3	0.71	Mudah
4	0.29	Sukar
5	0.43	Sedang
6	0.86	Easy
7	0.29	Hard
8	0.71	Easy
9	0.71	Easy
10	0.57	Medium
11	0.57	Medium
12	0.71	Easy
13	0.86	Medium
14	0.29	Hard
15	0.57	Medium

Based on the difficulty level test table on the Package B questions, the results obtained are 6 items fall into the easy category, 6 items in the medium category, and 3 items in the difficult category.

After testing the difficulty level of each question item, the discriminatory power test was carried out on the questions of packages A and B to determine the characteristics of the instrument questions made. The results of the discriminatory power test for package A can be seen in the following Table 9:

Table 9. Discriminatory Power Test of Question Package A

Discriminatory Power Test						
Question Number	Upper Average	Group	Lower Average	Group	Discriminatory Power	Criteria
1	1.00		0.50		0.50	Good
2	1.00		0.25		0.75	Very Good
3	0.67		0.25		0.42	Good
4	0.67		0.00		0.67	Good
5	1.00		0.25		0.75	Very Good
6	1.00		0.25		0.75	Very Good
7	0.67		0.00		0.67	Good
8	1.00		0.50		0.50	Good
9	1.00		0.50		0.50	Good
10	1.00		0.00		1.00	Very Good
11	1.00		0.00		1.00	Very Good
12	0.67		0.00		0.67	Good
13	1.00		0.50		0.50	Good
14	1.00		0.25		0.75	Very Good
15	1.00		0.25		0.75	Very Good

Based on the discriminatory power test table for Package A, it was found that 8 items falls into the good criteria in distinguishing students from the upper and lower groups, and 7 items had very good criteria in distinguishing

students from the upper and lower groups. Then the results of the discriminatory power test on Package B questions can be seen in the following Table 10:

Table 10. Discriminatory Power Test for Question Package B

Discriminatory Power Test						
Question Number	Upper Average	Group	Lower Average	Group	Discriminatory Power	Criteria
1	1.00		0.50		0.50	Good
2	1.00		0.00		1.00	Very Good
3	1.00		0.50		0.50	Good
4	0.67		0.00		0.67	Good
5	1.00		0.00		1.00	Very Good
6	1.00		0.75		0.25	Fair
7	0.67		0.00		0.67	Good
8	1.00		0.50		0.50	Good
9	1.00		0.50		0.50	Good
10	1.00		0.25		0.75	Very Good
11	1.00		0.25		0.75	Very Good
12	1.00		0.50		0.50	Good
13	1.00		0.75		0.25	Fair
14	0.67		0.00		0.67	Good
15	1.00		0.25		0.75	Very Good

Based on the discriminatory power test table for Package B, it was found that 5 items

had very good criteria in distinguishing upper and lower group students, then 8 items had

good criteria in distinguishing upper and lower group students, and 2 items had criteria are fairly sufficient to distinguish students from the upper and lower groups.

The Equality of Ability of Slow Learners with Special Needs with Normal Students in Inclusive Elementary Schools

According to Lee, et al. (2001), the purpose of equalizing test scores is to have a score that can be compared. As explained by Dorans (2004: 228), a score can be compared with another score, if both measure the same characteristics and are expressed in the same metric. Therefore, even though the two scores measure the same characteristics, if the scales are different, the two scores cannot be compared.

In the case of Package A and Package B, equalization is carried out to determine the equivalence of the ability of Slow learners with special needs with the ability of normal students. Equalization uses the classical test theory of the mean equalizing method which is suitable for small sample sizes (Babcock, 2012: 7). The mean equalizing formula is entered in the excel formula and resulted the following data:

Table 11. Equalization using the Mean Equating Method

	Y	X	my(x)	Value
	15	0	0.86	6
	15	11	11.86	79
	12	1	1.86	12
	9	15	15.86	106
	1	15	15.86	106
	1	3	3.86	26
	8	10	10.86	72
Mean	8.71	7.86	8.71	58.14

In Table 11, it is known that the classical test theory uses the "mean equating" method of equalization. The mean method is appropriate for use in equivalence tests with a small number of samples. According to Sha (2017: 9), the mean method is applied to the equating statistical model of CTT items, used

for equations with observed scores from the old form and the new form. The new shape is taken as the base shape, the old shape is equated with the new shape to get an equal cut score of the new shape. Equalization using the mean method shows the score x on form X which is the total score of each student in Package A is transformed to a scale of form Y which is the total score of each student on Package B produces the same mean value as the mean on the form Y scale, which is 8.71. Package A has a cognitive level below Package B, so it can be concluded that the results of the equalization of the two question packages indicate that the average score of students will increase if converted to an ability level above it. Students who have low scores and high scores have the same conversion direction, i.e. each score obtained will be greater.

CONCLUSION

Based on this research's results and discussion, it can be concluded that the assessment instrument developed focuses on evaluating the learning outcomes of slow learners with special needs in inclusive schools. There are 2 package questions made, namely package A which contains the lower level cognitive level (LOTS) which is used for learners with special needs, and package B which contains the upper level cognitive level (HOTS) which is used for normal students.

The results of the instrument validation test by 4 expert validators have met content validity as evidenced by the acquisition of an agreement index (Aiken index) which is in the range of 0.80 to 1.00. Inter-expert reliability was analyzed using Intraclass Correlation Coefficients (ICC) and obtained a result of 0.796 with good category. The reliability of large-scale test instruments is in the very reliable category, seen from the Kuder Richardson (KR) 20 Package A value of 0.983 and Package B value of 0.976.

The characteristics of the instrument were analyzed by calculating the difficulty level and the differentiating power of the

questions. Package A and Package B were dominated by items with moderate difficulty level. The discriminatory power of Package A questions is in the very good and good category in distinguishing students from the upper and lower groups. While Package B is in the very good, good, and fair category in distinguishing students from the upper and lower groups. Then for the results of equalizing the ability of slow learners with special needs with normal students using the mean equating method, it shows that the average score of students will increase if it is converted to a higher ability level.

REFERENCES

- Abdurrahman, M. (2010). *Pendidikan Bagi Anak Berkesulitan Belajar*. Jakarta: Rineka Cipta.
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131-142.
- Akbar, S. (2013). *Instrumen Perangkat Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Alfianto, F., Sumaryanto, T., & Utomo, U. (2015). Pengembangan Instrumen Penilaian Apresiasi Seni Musik Materi Seni Budaya Sekolah Menengah Pertama. *Journal Of Educational Research And Evaluation Vol. 2*, 2015 ISSN 2252 - 6420.
- Arifin, Z. (2014). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Babcock, B., Albano, A., & Raymond, M. (2012). Nominal weights mean equating: A method for very small samples. *Educational and Psychological Measurement*, 72(4), 608-628.
- Chineye, P. (2012). Evaluation and Assessment Frameworks for Improving School Outcomes Common Policy Challenges. *ACADEMIA Accelerating the Worlds Research*, 1-10.
- Dorans, N. J. (2004). Equating, concordance and expectation. *Jurnal Applied Psychological Measurement*, 28(4), 219-226
- Garnida, D. (2015). *Pengantar Pendidikan Inklusif*. Bandung: Refika Aditama.
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate IBM SPSS 23*. Semarang: Badan Penerbit Universitas Diponegoro.
- Haryono, H., Ahmad, S., & Sri, W. (2015). Evaluasi Pendidikan Inklusif Bagi Anak Berkebutuhan Khusus (ABK) di Provinsi Jawa Tengah. *Jurnal Penelitian Pendidikan*, 32(2) 2015, 119-126.
- Haug, P. (2017). Understanding Inclusive Education: Ideals and Reality. *Scandinavian Journal of Disability Research*, 19(3), 206-217.
- Kass, R. A. & Tinsley, H. E. A. (2009). Factor Analysis. *International Journal of Leisure Research*, 11, 120-138.
- Kim, S-H., & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the grade response model. *Applied Psychological Measurement*, 26, 25-41.
- Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, 25(4), 357-372.
- Malak, S. (2013). Inclusive Education Reform in Bangladesh: Pre Service Teachers' Responses to Include Students with Special Educational Needs in Regular Classrooms. *International Journal of Instruction*, 6(1), 195-214.
- Mehta, S., Bastero-Caballero, R. F., Sun, Y., Zhu, R., Murphy, D. K., Hardas, B., & Koch, G. (2018). Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. *Statistics in medicine*, 37(18), 2734-2752.
- Miller, W., Collings, J., & Kneale, P. (2015). Inclusive Assessment. *PedRIO with Plymouth University*, 7(7), 1-63.
- Ogasawara, H. (2001). Item response theory true score equatings and their standard errors. *Jurnal of Educational and Behavioral Statistic*, 26, 31-50.
- Pramesty, N. (2017). *Analisis Pelaksanaan Penilaian Proses dan Hasil Pembelajaran pada Sekolah Inklusif di SD Negeri Slerok 2 Kota Tegal*. Skripsi: Universitas Negeri Semarang.
- Pratiwiningtyas, B. N., Susilaningih, E., & Sudana, I. M. (2017). Pengembangan Instrumen Penilaian Kognitif untuk Mengukur Literasi Membaca Bahasa Indonesia Berbasis Model Pirls pada Siswa Kelas IV SD. *Journal of Research and Educational Research Evaluation*, 6(1), 01-09.
- Qomari, R. (2008). Pengembangan Instrumen Evaluasi Domain Afektif. *Jurnal Pemikiran Pendidikan INSANIA*, 13(1), 2008.

- Salomi D.S., & Sundaram, A.M. (2018). Slow Learning - Causes, Problems and Solutions. *International Journal of Management, Technology and Engineering*, 8(XII), 4209 – 4214.
- Sangeeta, M. (2009). Effect of Intervention Training on Mental Abilities of Slow Learners: *International Journal Education Science*, 1(1), 61-64.
- Schuelka, M, J. (2018). Implementing Inclusive Education. *Anthropology and Education Quarterly*, 49(2), 183–200..
- Sha, S. (2017). The Feasibility of Pre-equating with Classical Test Theory under Equivalent Group Design. *Center for Educational Measurement Excelsior College Albany, NY, 12203; Presented at the 48th Annual meeting of the Northeastern Education Research Association, Trumbull, CT, 2017*.
- Shaw, S.R., Grimes, D., & Bulman, J. (2005). Educating slow learners: are charter schools the last, best hope for their educational success? *American Journal of Mental Retardation*, 5(1), 463-469.
- Soleh, A., Khumaedi, M., & Pramono, S. E. (2017). Pengembangan Instrumen Penilaian Mata Pelajaran PKn Standar Kompetensi Memahami Kedaulatan Rakyat Dalam Sistem Pemerintahan di Indonesia. *Journal of Research and Educational Research Evaluation*, 6(1), 71-80.
- Sukmadinata, Nana Syaodih. (2017). *Metode Penelitian Pendidikan*. Bandung : PT. Remaja Rosdakarya
- Wells, C.S., & Wollack, J.A. (2003). *An Instructor's Guide To Understanding Test Reliability, Paper, Testing & Evaluation Services*, University of Wisconsin
- Schuelka, M.J. (2018). Implementing Inclusive Education. *Anthropology and Education Quarterly*, 49(2), 183–200.
- Yusuf, M., Choiri, S., & Supratiwi, M. (2017). Evaluation of Inclusive Education Policies at the Level of Primary and Secondary Education in Indonesia. *Jurnal Penelitian Dan Pengembangan Pendidikan Luar Biasa*, 4(2), 147-154.
- Zuliani, D., Florentinus, T. S., & Ridlo, S. (2017). Pengembangan Instrumen Penilaian Karakter pada Siswa Kelas IV Sekolah Dasar. *Journal of Research and Educational Research Evaluation*, 6(1), 46-54.