



The Concept of Heat Transfer measured by Cognitive Domain Assessment Instruments

* N K Yuniasih^{1*}, K Yudiana², I G N Japa³

^{1,2,3} Program Studi Pendidikan Guru Sekolah Dasar, Universitas Pendidikan Ganesha, Indonesia

ARTICLE INFO

Article history:

Received in revised form
05 January 2021
Accepted 27 January 2021
Available online 25
February 2021

Kata Kunci:
Konsep Perpindahan Kalor,
Ranah Kognitif

Keywords:
Concept Of Heat Transfer;
Domain Assessment

ABSTRAK

Kurangnya ketersediaan instrumen penilaian yang digunakan serta tingkat kognitif C3-C6 yang belum terpenuhi mengakibatkan penilaian yang dilakukan kurang berkualitas. Penelitian ini bertujuan untuk mengembangkan instrumen penilaian ranah kognitif pada materi konsep perpindahan kalor kelas V SD. Penelitian ini merupakan penelitian pengembangan yang menggunakan model 4D (four-D models) dengan tahap pengembangan yang terdiri dari define, design, develo, dan disseminate. Subjek dari penelitian ini adalah instrumen penilaian ranah kognitif dengan objek penelitian yaitu kualitas instrumen. Pengukuran validitas isi dilakukan pada 2 orang ahli/judges dengan menggunakan metode nontes dan instrumen lembar validitas ahli. Pengukuran kualitas instrumen dilakukan dengan uji coba lapangan sebanyak 106 siswa menggunakan instrumen tes objektif pilihan ganda.

Data validitas isi yang diperoleh dianalisis dengan rumus Gregory, sedangkan data hasil uji coba lapangan dianalisis validitas butir, reliabilitas, daya beda, dan tingkat kesukarannya. Hasil penelitian menunjukkan bahwa validitas isi yang diperoleh sebesar 0,97 (sangat tinggi) dengan hasil uji validitas butir diperoleh 30 butir soal dinyatakan valid. Indeks reliabilitas instrumen yang dikembangkan sebesar 0,85 (sangat tinggi). Rata-rata koefisien daya beda yang diperoleh sebesar 0,51 (baik) dengan tingkat kesukaran perangkat tes sebesar 0,52 (sedang). Hasil tersebut menunjukkan instrumen penilaian ranah kognitif yang dikembangkan layak dan dapat digunakan dalam melakukan penilaian pembelajaran pada ranah kognitif.

ABSTRACT

The lack of availability in assessment instruments used and unmet cognitive levels of C3-C6 resulted in poor quality assessments. This study aims to develop cognitive domain assessment instruments on the concept of heat transfer material in grade V elementary school. This research was development research that used 4D models (Four-D models) to define, design, develop, and disseminate. The subject of study was the assessment instrument of the cognitive realm with the object of research, that was the instrument quality. Measurement of the validity of the content was done on two experts/judges using the non-test method and instruments of the expert validity sheet. Instrument quality measurement was conducted with field trials of 106 students using multiple-choice objective test instruments. The content validity data obtained were analyzed with the Gregory formula, while the field test result data analyzed the validity of the grain, reliability, different power, and difficulty level. The results showed that the validity of the content obtained by 0.97 (very high) with the test results of the item's validity got 30 points of the question was declared valid. The instrument reliability index was developed by 0.85 (very high). The average coefficient of different power obtained is 0.51 (good) with a test device difficulty of 0.52 (medium). The result demonstrated cognitive domain assessment instruments that were developed feasible and can be used in conducting learning assessments in the cognitive realm.

1. Introduction

The learning process conducted in the school cannot be separated from the assessment. In fact, an assessment was a process to collect various data on the development of learning and student learning outcomes so that it can be meaningful information in competency achievement decision making (Asiah et al., 2017; Safitri & Oktavia, 2017). Assessment can be used to determine the development of education

obtained by students in their lives (Sonmes et al., 2021). The implementation of the assessment should be adjusted to what will be measured and use the appropriate measuring instruments. Measuring instruments that can be used in carrying out assessments are instruments. Instruments can be used to collect data on learning outcomes and evaluation of the learning process that has been done (Desilva et al., 2020; Gaol et al., 2017). Good instruments to be used in conducting assessments must be adjusted to the appropriate instrument procurement procedures and pay attention to students' level of thinking ability.

However, based on the results of interviews with grade V teachers at SD Negeri 2 Kalibukbuk on Tuesday, October 27, 2020; it was obtained information that in the assessment of the cognitive realm, both used in daily tests, midterm exams, and final semester exams are attempted to match the cognitive level, but in the making of questions at higher cognitive level teachers still have difficulties. In addition, the manufacture is only done in the form of questions and not yet equipped with instruments. Based on the analysis results of the assessment of the midterm science exam content on theme 1 and theme 2 grade V lesson year 2020/2021 at SD Negeri 2 Kalibukbuk, obtained the fact that the question used can only measure cognitive levels in C1 and C2. As for the cognitive level of C3-C6, especially in science, content can not be fulfilled. Theme instrument 1 on the science content consisting of 2 description questions shows that both questions are at the cognitive level of C2. Meanwhile, the theme instrument 2 on the science content consisting of 2 questions of description shows that both questions are at the cognitive level of C1. Teachers often have difficulty assessing learning outcomes (cognitive) caused by the evaluation tools used have not been able to measure properly and have not been following what should be measured (Mudanta et al., 2020; Wirayasa et al., 2020). Teachers' problem is also related to the availability of cognitive assessment instruments to measure high levels of thinking ability (Arifin & Retnawati, 2017). In addition, the questions used by teachers in carrying out assessments have never been tested validity level, so it can not be known eligibility (Nugroho & Airlan, 2020).

Judging from the problems found, if no improvement is made or solutions to overcome it, the assessment process is not good, so the learning is less qualified. An alternative solution that can be done is to develop an assessment instrument. This is in line with some research that has been conducted before. The development of self-assessment instruments obtains results in the valid and reliable category (Yustiana & Ulia, 2019). Cognitive assessment instruments that have been developed are in the category worth looking at from the suitability of the validity of the grain, reliability, difficulty level, and power difference so that it is very well used by teachers (Mustari, 2016; Pratiwiningtyas et al., 2017; Utami & Wardani, 2020). Implementation of the development of instruments value of civic subject obtained the results of the development of instruments that are declared validity and reliability has met the criteria (Soleh et al., 2017). In addition, *higher-order thinking skills* for science material developed obtained results that were in the category of feasible (Taufiqurrahman et al., 2018). Furthermore, the development of process assessment instruments to measure the efficacy of science and student activity obtains valid and reliable results so that it is feasible to use (Arini et al., 2017). Development is also carried out on the assessment instrument of science learning results that can be declared valid and reliable. It is feasible to be used in assessing the results of science learning (Pujawan et al., 2020; Widyaningsih et al., 2020). Also, the development and validation of knowledge assessment instruments also obtains valid and reliable results (Bala et al., 2020). Judging from the research of instrument development that has been done before can be known that in general the instruments that have been developed before are mostly only tested validity and reliability only, while for different power and difficulty levels are still not noticed. In addition, the development of instruments carried out at the elementary level, especially in the cognitive realm is still very limited.

Judging from the problems and solutions that have been done, an alternative is needed to overcome the difficulties and shortcomings of the previous solution. Therefore, cognitive domain assessment instruments in the concept of heat transfer material in grade V elementary school. The development of this instrument is important to obtain an instrument that is capable and feasible to be used in conducting quality cognitive field assessments. Cognitive domain assessment instruments developed have a newness that can distinguish them from previously developed instruments. The difference lies in the material developed in the instrument that is the concept of heat transfer. In addition, the cognitive level achieved in the instruments developed from C3-C6 has led to a high level of thinking ability and is adapted to the needs of the field. The advantages of the instrument developed can measure the ability to think at a high level, has been following the curriculum used that is the curriculum 2013, and the material in the instrument is adapted to students' daily lives. The development of cognitive domain assessment instruments aims to develop cognitive domain assessment instruments on the concept of the heat transfer material on grade V elementary school. The implementation of the development of this

instrument is the availability of cognitive domain assessment instruments with calorific displacement concept materials that are eligible to be used in measuring or assessing the learning process conducted.

2. Method

This type of research was developmental research that developed assessment instruments in the cognitive realm with a 4D development model (*Four-D model*). Thiagarajan (in Mulyatiningsih, 2014) revealed that the 4D development model consists of defining, designing, developing, and disseminate stages. At the defining stage, identifying the needs and collection of field information related to the product needs to be developed. At the design stage, the design of cognitive domain assessment instrument device on Science content. At the development stage, the instrument's development and revision through expert trials /judges and field trials. Meanwhile, the dissemination of products at the disseminate stage was developed had been tested so that it was feasible to use. The subject of study was the instrument of assessment of the cognitive realm with the object of research, namely the quality of the instrument. The quality intended in this study consisted of the validity of the content, the validity of the grain, reliability, different power, and the difficulty level of the instrument.

Based on the problem studied, namely the assessment of the cognitive realm on science content, the method of data collection used in this study is a non-test method to obtain the results of the content validity test of 2 judges by providing relevant or irrelevant assessments on the expert validity sheet. In addition, test methods were also used to obtain field test results data by conducting tests on 106 students in multiple-choice objective tests. Analysis of the data obtained in this study includes analysis of validity, reliability, power differences, and difficulty levels of instruments. Instrumentality is essential for perfecting instruments in the accurate decision-making of something to measure (Blazquez et al., 2017; Morad et al., 2021; Vial et al., 2021). The instrument validity analysis consisted of a content validity test and an item validity test. The examination of the validity of the content was conducted using cross-tabulation 2x2, and Gregory formula that was through expert assessment (judges) following field, in this study experts or experts come from the faculty of education, the Ganesha University of Education who had expertise in the field of science. The validity value of the content obtained reflects the entire test item tested. The validity value of the content obtained can then be seen based on the content validity coefficient in the following table.

Table 1. Content Validity Coefficient Category

Coefficient	Validity
0,80 – 1,00	Very High
0,60 – 0,79	High
0,40 – 0,59	Moderate
0,20 – 0,39	Low
0,00 – 0,19	Very Low

(Arikunto, 2018)

The validity of an item is a measuring instrument that can be used to obtain the validity or absence of each question item in the device. To test the validity of the formula item used was a correlation of biserial points (r_{pbi}) because the test item was in the form of multiple-choice objectives or so-called dichotomies that in the scoring give a score of 1 if the answer was correct and a score of 0 if the answer is wrong. Test items can be declared valid if the calculated r is greater than the table r with a significance level or a 5% error level. Reliability was used to measure the level of skill of an instrument. Measurement in question was when the instrument was used to measure the same conditions will produce results that are not much different. Reliability is the consistency of instruments that provides the same information when tested repeatedly (Casanova et al., 2021; Ko et al., 2017; Sa'idah et al., 2019). The formula used to perform reliability tests on instruments with dichotomy properties is the Kudor Richardson formula (KR-20). The reliability coefficient obtained is then compared to the criteria in the following Table 2.

Table 2. Instrument Reliability Criteria

Interval	Criteria
0,81 – 1,00	Very High
0,61 – 0,80	High
0,41 – 0,60	Moderate
0,21 – 0,40	Low
0,00 – 0,20	Very Low

(Akhsan et al., 2019)

The different power of an instrument was the ability of the instrument to distinguish between students who were classified as smart and students who were classified as less smart. In analyzing the different instruments, a rank classification was performed to determine the upper group of students and the lower group of students. The practical ranking of sciences obtained the upper group and bottom group based on the score obtained, then 27% of the students who obtained the highest score were taken as the top group, and 27% of the students who obtained the lowest score were taken as the bottom group (Candiasa, 2011). Suppose the "D" obtained is negative, the problem very bad, and must be discarded (Koyan, 2011). The result of the calculation of different power obtained was then compared with the criteria in the following table.

Table 3. Difference Criteria (D)

Interval	Criteria
0,71 – 1,00	Very Good
0,40 – 0,70	Good
0,20 – 0,39	Enough
0,00 – 0,19	Bad

(Koyan, 2011)

The difficulty level was difficult or not a test that can be known through the ability of the test taker to give the correct answer. The degree of difficulty was carried out by testing against the difficulty level of test grains and the degree of difficulty of the test device. The difficulty level of the item is the proportion of correct answers students in one question (Ndiung & Jediut, 2020). Meanwhile, the difficulty level of the test device was the average proportion of students' correct answers in the overall test device. The difficulty level criteria could be seen in the following table.

Table 4. Difficulty Level Criteria

Interval	Criteria
0,00 – 0,29	Hard
0,30 – 0,70	Moderate
0,71 – 1,00	Easy

(Koyan, 2011)

3. Result and Discussion

In define obtained information based on interviews conducted with grade V teachers at SD Negeri 2 Kalibukbuk on Tuesday, October 27, 2020, that in the preparation of cognitive field assessment questions, both used in daily tests, midterm exams, and final semester exams conducted were attempted to match the cognitive level. However, in the preparation of tests at a higher cognitive level, teachers still have difficulties. In addition, information was obtained that the curriculum used was the 2013 curriculum. Meanwhile, based on the analysis of instruments used in the midterm exams science content on theme 1 and theme 2 class V lesson year 2020/2021 at SD Negeri 2 Kalibukbuk obtained the fact that the instrument used could only measure the cognitive level of C1-C2 while the cognitive realm of C3-C6 can not be fulfilled. The results of the analysis obtained can be seen in table 5.

Table 5. Instrument Analysis Results

No.	Cognitive Level	Theme 1 (Item Number)	Theme 2 (Item Number)	Number of Item
1.	Remember (C1)	-	5 and 9	2
2.	Understand (C2)	5 and 10	-	2
3.	Applying (C3)	-	-	0
4.	Analyze(C4)	-	-	0
5.	Evaluate (C5)	-	-	0
6.	Create (C6)	-	-	0
Total				4

At the design stage, the results of basic competency analysis (KD) on the science content in this study were limited to the theme of 6 temperatures and displacement with basic competency 3.6 applying

the concept of heat transfer in daily life. Based on basic competencies, the preparation of achievement indicators as many as 12 indicators. Then, the initial arrangement of the grid with the design can be seen in the following table.

Table 6. Cognitive Domain Assessment Instrument Grid

No.	Basic Competency	Indicator	Cognitive Level	Dimensions	Item Number	Total
1.	3.6 Applying the concept of heat transfer in daily life.	3.6.1 Determining the source of heat energy in daily life.	C3	Factual	1 2	2
		3.6.2 Applying the concept of heat transfer in daily life.	C3	Conceptual	3 4	2
		3.6.3 Determining the concept of heat transfer in daily life.	C3	Conceptual	5 6	2
		3.6.4 Investigating conduction heat transfer in daily lives.	C3	Conceptual	7 8	2
		3.6.5 Investigating convection heat transfer in daily lives.	C3	Conceptual	9 10 11	3
		3.6.6 Analyzing radiation heat transfer in daily lives.	C4	Conceptual	12 13 14	3
		3.6.7 Grouping objects that can accelerate heat transfer in daily lives.	C4	Metacognitive	15 16 17 18	4
		3.6.8 Finding objects that can inhibit heat transfer in daily lives.	C4	Metacognitive	19 20 21 22	4
		3.6.9 Summing up the concept of heat transfer in daily lives.	C5	Conceptual	23 24 25	3
		3.6.10 Detailing the concept of heat transfer in daily lives.	C5	Conceptual	26 27 28 29	4
		3.6.11 Categorizing conductors and insulators in daily lives.	C6	Conceptual	30 31 32	3
		3.6.12 Develop a way of utilizing the properties of conductors and insulators in daily lives.	C6	Procedural	33 34 35	3
Total						35

After the initial preparation of the grid was carried out the initial preparation of instruments designed as many as 35 points of questions with an arrangement including (1) the title of the instrument, (2) the education unit, (3) the class/semester, (4) the theme, (5) the type of question, (6) the allocation of time, (7) the number of questions, (8) instructions, and (9) the question. At the development stage, development was carried out with expert tests/judges and field trials. Expert tests/judges were conducted on two experts from the faculty of education, the Ganesha University of Education, to obtain data on the validity of the contents of the instrument. The results of expert assessment/judges brought could be seen in table 7.

Tabel 7. Expert Judgement Results

Judges 1		Judges 2	
Relevant	Not Relevant	Relevant	Not Relevant
1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35.	10.	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35.	-

The expert assessment/judges data were then included in cross×2 tabulation and analyzed using the Gregory formula. Based on the calculation result, the instrument content validity index was 0.97. The calculation results showed that the cognitive domain assessment instruments on science content tested had a very high category of content validity. Field trials were conducted on grade VI students in 4 different elementary schools, namely SD Negeri 1 Kalibukbuk as many as 13 students, SD Negeri 2 Kalibukbuk as many as 38 students, SD Negeri 6 Banyuning as many as 21 students, and SD Negeri 1 Landih as many as 34 students. So the number of respondents was 106 students. Field test result data was used to analyze item validity, reliability, power differences, and difficulty levels. Item validity test data was analyzed using a coefficient formula of a biserial point correlation assisted by using a Microsoft Excel 2013 application. The results of the item validity test were presented in table 8.

Tabel 8. Results of Item Validity Test

Item Validity	Item Number	Total	Percentage
Valid	1, 3, 5, 6, 7, 8, 9, 11, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 30, 32, 32, 33, 34.	30	88,24%
In	2, 4, 25, 29	4	11,76%

Based on the item validity test results, out of the 34 question items, there were 30 questions declared valid and four questions declared invalid. The results showed that the percentage of questions said valid was 88.24%, and the percentage of questions said invalid was 11.76%. A declared invalid can indicate the quality of the problem was not good, so the invalid question was declared void and not used as an assessment instrument after the item validity test was conducted reliability test on the question that had been declared valid. The instrument's reliability was analyzed using a formula that had dichotomy properties, namely the Kudor Richadson formula (KR-20) which was assisted by using a Microsoft Excel 2013 application. Based on the calculation results obtained, the instrument made had a reliability index of 0.85. The development of the calculation was compared with the criteria used as a reference to know that the instrument's reliability is very high. The calculation results showed that the instrument developed reliable so that it could be tested at any time with relatively similar results in equivalent respondents.

Different power tests were only performed on questions that were declared valid. Carrying out different power tests were carried out by determining the upper and lower groups of samples. The upper group of samples was obtained based on 27% of the total samples with the highest score. Meanwhile, the lower group of samples was obtained based on 27% of the total samples with the lowest score. Different power tests were calculated by using the help of Microsoft Excel 2013 applications. The results of different power tests were presented in table 9.

Tabel 9. Results of Different Power Test

Criteria	Item Number	Total	Percentage
Very Good	8, 18, 33,34	4	13,33%
Good	1, 3, 5, 6, 7, 9, 10, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 26, 27, 28, 30, 31, 32	23	76,67%
Enough	11, 22, 24	3	10%

Based on the results of different power calculations conducted on questions that were considered valid or 30 points of the question, obtained 3 points of the question had different power that was on the

criteria quite well, 23 points of the question had different power that was on the criteria well, and 4 points of the question had different power that was on the criteria very well. The average test of different power coefficients obtained based on the calculation of varying power grains was 0.51 with a good category. It shows that the instruments developed have a good ability to distinguish smart and less intelligent students or students who understand the material and students who did not understand the material. Difficulty level tests were only performed on valid questions. The difficulty level test was performed on each test item then continued with the difficulty level test. Calculation of difficulty levels was performed with the help of Microsoft Excel 2013 applications. The test results of the difficulty level of test items were presented in table 10.

Tabel 10. Results of Difficulty Level Test

Difficulty Level Category	Item Number	Total	Percentage
Hard	10, 12	2	6,67%
Moderate	3, 6, 7, 8, 9, 11, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 30, 31, 32, 33, 34	25	83,33%
Easy	1, 5, 15	3	10%

The analysis results showed that out of the 30 questions tested for difficulty, three questions were obtained with easy criteria, 25 questions with moderate criteria, and two questions with difficult criteria. After testing the test item's difficulty level, it was then continued by conducting a test of the difficulty level of the test device. The test device difficulty level index obtained was 0.52 with moderate criteria. Based on the calculation of the difficulty of the test item and the difficulty level of the test device, it can be interpreted that the instrument developed has good quality based on the degree of difficulty. At the dissemination stage, the dissemination is carried out by handing over the instrument in the form of hardcopy as much as 25 esksemplar and softcopy of the instrument to the teacher to be used in the learning activities carried out.

Discussion

The product development results showed that the instruments developed had been tested for validity, reliability, power differences, and difficulty levels. The validity and reliability of the instruments developed were very high, with good instrument differentiation and moderate difficulty levels that could be declared worthy of being used in assessing the cognitive realm in the learning process. The instruments developed can be considered worthy of judging by the appropriate curriculum used, namely the 2013 curriculum, the suitability of the materials developed, and the conformity with the characteristics of students. The instrument's feasibility was also supported by the advantages that have led to a high level of thinking ability with the cognitive realm of C3-C6 following the needs of the field. A high level of thinking in the learning and teaching process is an important aspect of student development (Nurhayati & Angraeni, 2017). In addition, the requirements of a good instrument at least meet the validity and reliability of the instrument and the difficulty and power of differences that need to be considered (Iskandar & Rizal, 2017; Santee et al., 2019).

The results were in line with some of the findings from previous studies. The development of cognitive assessment instruments to measure reading literacy skills is eligible to be used to conduct an assessment (Pratiwiningtyas et al., 2017). The thematic learning cognitive assessment instrument developed produces a good and viable product because it can measure students' cognitive abilities (Utami & Wardani, 2020; Widiana et al., 2020). In addition, research on the reliability and validity of instruments obtained valid and reliable results with a content validity index of 0.97 with a validity ratio of 0.83 and reliability $r = 1/4$ 0.85, $p < 0.001$ (Shirali et al., 2018). The development of cognitive instruments in the subject of static fluids produces excellent tools for use by teachers in student competency assessment (Mustari, 2016). Also, the assessment instrument to measure the learning motivation developed makes instruments in the category of feasible, has a very high validity value of 0.85, and a high degree of reliability of 0.80 (Krismony et al., 2020). Based on previous research, the instrument developed in this study had an update or difference in the material produced in the instrument, namely the concept of heat transfer that had never been developed in previous research. In addition, the cognitive level achieved in the instruments developed from C3-C6 had led to a high level of thinking ability and was adapted to the needs of the field. The results of this study have implications for the availability of cognitive domain assessment instruments science content that was feasible because it had been tested for quality. The quality in question consists of validity, reliability, different power, and difficulty level to take measurements or assessments in the learning process conducted.

4. Conclusion

The research and discussion results show that the cognitive domain assessment instruments developed are valid and reliable and have been tested for different power and difficulty levels. Thus, the instruments developed will give results that are not much different when tested under the same conditions and distinguish students who were classified as smart and less smart.

References

- Akhsan, H., Wiyono, K., Novianti, R., Melvany, N. E., & Ariska, M. (2019). Pengembangan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Materi Fluida dan Getaran Harmonis. *Jurnal Inovasi Pendidikan*, 9(2), 33–40. <http://sij-inovpend.ejournal.unsri.ac.id/index.php/sij-inovpend/index>.
- Arifin, Z., & Retnawati, H. (2017). Pengembangan Instrumen Pengukur Higher Order Thinking Skills Matematika Siswa SMA Kelas X. *PYTHAGORAS: Jurnal Pendidikan Matematika*, 12(1), 98–108. <https://doi.org/10.21831/pg.v12i1.14058>.
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Arini, F. Z. R., Susilaningsih, E., & Dewi, N. K. (2017). Pengembangan Instrumen penilaian Proses untuk Mengukur Keterampilan Sains dan Aktivitas Siswa SMP. *Journal of Innovative Science Education*, 6(2), 170–178. <https://doi.org/10.15294/jise.v6i2.19715>.
- Asiah, H. A., Susilaningsih, E., & Nuswowati, M. (2017). Inovasi Model Penilaian Proses Pada Pembelajaran Kimia Untuk Mengukur Keterampilan Laboratorium Dan Aktivitas Siswa. *Jurnal Inovasi Pendidikan Kimia*, 11(2), 1997–2007. <https://journal.unnes.ac.id/nju/index.php/JIPK/article/view/10624>.
- Bala, A. A., Jatau, A. I., Yunusa, I., Mohammed, M., Mohammed, A. K. H., Isa, A. M., Sadiq, W. A., Gulma, K. A., Bello, I., Malami, S., Michael, G. C., & Chedi, B. A. Z. (2020). Development and validation of antsnake venom knowledge assessment tool (AKAT) for healthcare practitioners. *Toxicon: X*, 8(September), 100064. <https://doi.org/10.1016/j.toxcx.2020.100064>.
- Blazquez, R. A., Corchon, S., & Ferrer Ferrandiz, E. (2017). Validity of instruments for measuring the satisfaction of a woman and her partner with care received during labour and childbirth: Systematic review. *Midwifery*, 55(June), 103–112. <https://doi.org/10.1016/j.midw.2017.09.014>.
- Candiasa, M. (2011). *Pengujian Instrumen Penelitian Disertai Aplikasi ITEMAN dan BIGSTEPS*. Unit Penerbitan Universitas Pendidikan Ganesha.
- Casanova, J. R., Gomes, C. M. A., Bernardo, A. B., Núñez, J. C., & Almeida, L. S. (2021). Dimensionality and reliability of a screening instrument for students at-risk of dropping out from Higher Education. *Studies in Educational Evaluation*, 68(1–8). <https://doi.org/10.1016/j.stueduc.2020.100957>.
- Desilva, D., Sakti, I., & Medriati, R. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Fisika Berorientasi HOTS (High Order Thinking Skills) pada Materi Elastisitas dan Hukum Hooke. *Jurnal Kumparan Fisika*, 3(1), 41–50. <https://doi.org/https://doi.org/10.33369/jkf.3.1.41-50>.
- Gaol, P. L., Khumaedi, M., & Masrukan. (2017). Pengembangan Instrumen Penilaian Karakter Percaya Diri pada Mata Pelajaran Matematika Sekolah Menengah Pertama. *Jurnal of Educational Research and Evaluation*, 6(1), 63–70. <https://doi.org/https://doi.org/10.15294/jrer.v6i1.16209>.
- Iskandar, A., & Rizal, M. (2017). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 12–23. <https://doi.org/https://doi.org/10.21831/pep.v22i1.15609>.
- Ko, M. M., Lee, M. S., Birch, S., & Lee, J. A. (2017). The reliability and validity of instruments measuring pattern identification in Korean medicine: A systematic review. *European Journal of Integrative Medicine*, 15(May), 47–63. <https://doi.org/10.1016/j.eujim.2017.09.005>.
- Koyan, I. W. (2011). *Asesmen dalam Pendidikan*. Singaraja: Universitas Pendidikan Ganesha Press.
- Krismony, N. P. A., Parmiti, D. P., & Japa, I. G. N. (2020). Pengembangan Instrumen Penilaian Untuk Mengukur Motivasi Belajar Siswa SD. *Jurnal Ilmiah Pendidikan Profesi Guru*, 3(2), 249–257. <https://doi.org/http://dx.doi.org/10.23887/jippg.v3i2>.
- Morad, S., Ragonis, N., & Barak, M. (2021). The validity and reliability of a tool for measuring educational innovative thinking competencies. *Teaching and Teacher Education*, 97, 2–11. <https://doi.org/10.1016/j.tate.2020.103193>.
- Mudanta, K. A., Astawan, I. G., & Jayanta, I. N. L. (2020). Instrumen Penilaian Motivasi Belajar dan Hasil Belajar IPA Siswa Kelas V Sekolah Dasar. *Jurnal Mimbar Ilmu*, 25(2), 262–270. <https://doi.org/http://dx.doi.org/10.23887/mi.v25i2.26611>.
- Mulyatiningsih, E. (2014). *Metode Penelitian Terapan Bidang Pendidikan*. Bandung: Alfabeta.
- Mustari, M. (2016). Pengembangan Instrumen Ranah Kognitif pada Pokok Bahasan Fluida Statis SMA/MA. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 5(1), 121–130.

- <https://doi.org/10.24042/jpifalbiruni.v5i1.112>.
- Ndiung, S., & Jediut, M. (2020). Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, 10(1), 94–111. <https://doi.org/10.25273/pe.v10i1.6274>.
- Nugroho, A. N., & Airlan, G. S. (2020). Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Pembelajaran IPA Kelas 4 SD. *Jurnal Ilmiah Pendidikan Profesi Guru*, 3(3), 400–407. <https://doi.org/http://dx.doi.org/10.23887/jippg.v3i3>.
- Nurhayati, & Angraeni, L. (2017). Analisis Kemampuan Berpikir Tingkat Tinggi Mahasiswa (Higher Order Thinking) dalam Menyelesaikan Soal Konsep Optika melalui Model Problem Based Learning. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 3(2), 119–126. <https://doi.org/doi.org/10.21009/1.03201>.
- Pratiwiningsih, B. N., Susilaningsih, 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 Educational Research and Evaluation*, 6(1), 1–9. <https://doi.org/https://doi.org/10.15294/jrer.v6i1.16199>.
- Pujawan, I., Parmiti, D. P., & Astawan, I. G. (2020). Assessment Instrument of Social Attitude and Science Learning Outcomes of Grade IV Elementary School. *Jurnal Ilmiah Sekolah Dasar*, 4(4), 604–613. <https://ejournal.undiksha.ac.id/index.php/JISD/article/view/26580>.
- Sa'idah, N., Yulistianti, H. D., & Megawati, E. (2019). Analisis Instrumen Tes Higher Order Thinking Matematika SMP. *Jurnal Pendidikan Matematika*, 13(1), 41–54. <https://doi.org/10.22342/jpm.13.1.6619.41-54>.
- Safitri, D., & Oktavia, M. (2017). Implementasi Penilaian Autentik Kurikulum 2013 (Studi Kasus Guru IPS di SMP Labschool Jakarta). *Edukasi IPS*, 01(1), 31–40. <http://journal.unj.ac.id/unj/index.php/eips/article/view/3207>.
- Santee, J. A., Marszalek, J. M., & Hardinger, K. L. (2019). A critique of validity analysis from instruments assessing the affective domain. *Currents in Pharmacy Teaching and Learning*, 11(2), 218–229. <https://doi.org/10.1016/j.cptl.2018.11.010>.
- Shirali, G., Shekari, M., & Angali, K. A. (2018). Assessing Reliability and Validity of an Instrument for Measuring Resilience Safety Culture in Sociotechnical Systems. *Safety and Health at Work*, 9(3), 296–307. <https://doi.org/10.1016/j.shaw.2017.07.010>.
- 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. <https://doi.org/https://doi.org/10.15294/jrer.v6i1.16210>.
- Sonmes, F. D., Cuhadar, S., & Kahvecioglu, M. K. (2021). Successes, challenges, and next steps in implementing outcome-based assessment: The case of Istanbul Bilgi University Library. *Journal of Academic Librarianship*, 47(1), 1–8. <https://doi.org/10.1016/j.acalib.2020.102249>.
- Taufiqurrahman, Heryandi, M. T., & Junaidi. (2018). Pengembangan Instrumen Penilaian Higher Order Thinking Skill pada Mata Pelajaran Pendidikan Agama Islam. *Jurnal Pendidikan Administrasi Perkantoran*, 2(2), 199–206. <https://www.ojs.pps-ibrahimy.ac.id/index.php/jpii/article/view/74>
- Utami, D. A. P., & Wardani, N. S. (2020). Pengembangan Instrumen Penilaian Kognitif dalam Pembelajaran Tematik Kelas 5 SD. *LENTERA: Jurnal Ilmiah Kependidikan*, 13(1), 1–18. <http://stkipgribl.ac.id/jurnal/index.php/lentera/article/view/463>.
- Vial, A., van der Put, C., Stams, G. J. J. M., Dinkgreve, M., & Assink, M. (2021). Validation and further development of a risk assessment instrument for child welfare. *Child Abuse and Neglect*, 117(April), 1–11. <https://doi.org/10.1016/j.chiabu.2021.105047>.
- Widiana, I. G. K., Dantes, N., & Parmiti, D. P. (2020). Pengembangan Instrumen Penilaian Kognitif Pembelajaran Tematik Tema 8 Subtema 1 pada Siswa Kelas V SD. *Jurnal Ilmiah Pendidikan Profesi Guru*, 3(3), 478–485. <https://doi.org/http://dx.doi.org/10.23887/jippg.v3i3.29428>.
- Widyaningsih, I. A. T., Dibia, I. K., & Sudana, D. N. (2020). Instrument Appropriateness On The Assessment Of Fourth Grade Science Learning Outcomes In Elementary Schools. *Internasional Journal of Elementary Education*, 4(4), 472–480. <https://ejournal.undiksha.ac.id/index.php/IJEE/article/download/27209/17627>.
- Wirayasa, I. D. G. P., Darmayasa, I. P., & Satyawan, I. M. (2020). Pengembangan Instrumen Penilaian Hasil Belajar Ranah Kognitif Model 4D pada Materi Sepak Bola Berdasarkan Kurikulum 2013. *Jurnal Pendidikan Jasmani Olahraga Dan Kesehatan*, 8(3), 81–88. <https://ejournal.undiksha.ac.id/index.php/JJP/article/view/33760>.
- Yustiana, S., & Ulia, N. (2019). Pengembangan Instrumen Penilaian Diri Berbasis Pembelajaran Kontekstual. *Profesi Pendidikan Dasar*, 6(2), 180–188. <https://doi.org/10.23917/ppd.v1i2.8998>.