Development of Diagnostic Test Instruments as Placement Tests to Determine Students' Cognitive Learning Levels on Acid-Base Material

Saifatun Nur Hafidzah Z, Rinaningsih

Chemistry Education Study Program, Faculty of Mathematics and Natural Science,
Universitas Negeri Surabaya, Indonesia
*Email: fydasaaifatun@gmail.com

ABSTRACT

This study aims to determine the validity and discriminatory power of diagnostic test instruments as placement tests on acid-base material and to determine the level of students' cognitive learning. This study was conducted in June - May of the 2024/2025 Academic Year. This diagnostic test instrument development research refers to the Research & Development (R&D) development method. Data collection in this study used tests, interviews, and questionnaires. The research stage begins with potential problems, data collection, product design stage, development of diagnostic test instruments and validation by experts, validity tests, difficulty level tests, and discriminatory power tests on small-scale tests and interpretation of students' cognitive levels on large-scale trials. The results of the assessment by expert validators showed that the criteria for the developed instrument were valid. The results of the analysis of the validity of the test items obtained 25 valid question items. The results of the test of the level of difficulty of the questions obtained 11 easy criteria questions and 14 medium criteria questions. The results of the discriminatory power test obtained 25 question items with very good discriminatory power. The results of the large-scale trial showed that the average interpretation of the results of the diagnostic test instrument was 78.81%. The percentage of students who passed the test was 70.3% while the percentage of students who failed was 30%. This shows that the diagnostic test instrument is effective in that it can improve students' understanding of acid-base chemistry material.

Keywords: *Development; Diagnostic Test Instruments; Cognitive Learning; Placement Test; Acid Base.*

ABSTRAK

Penelitian ini bertujuan untuk mengetahui validitas dan daya pembeda instrumen tes diagnostik sebagai tes penempatan pada materi asam basa dan untuk mengetahui tingkat kognitif belajar siswa Penelitian ini dilaksanakan pada Juni – Mei Tahun Ajaran 2024/2025. Penelitian pengembangan instrumen tes diagnostik ini mengacu pada metode pengembangan Research & Development (R&D). Pengumpulan data pada penelitian ini menggunakan tes, wawancara, dan kusioner. Tahap penelitian dimulai dengan potensi masalah, pengumpulan data, tahap perancangan produk, pengembangan instrumen tes diagnostik dan validasi oleh ahli, uji validitas, uji tingkat kesukaran, dan uji daya pembeda pada uji skala kecil serta interpretasi tingkat kognitif siswa pada uji coba skala luas. Hasil penilaian dari validator ahli menunjukkan kriteria instrumen yang dikembangkan valid. Hasil analisis validitas butir soal diperoleh 25 item soal valid. Hasil uji tingkat kesukaran soal diperoleh 11 item soal kriteria mudah dan 14 item soal kriteria sedang. Hasil uji daya pembeda diperoleh 25 item soal memiliki daya pembeda sangat bagus. Hasil uji coba skala luas menunjukkan rata - rata interpretasi hasil instrumen tes diagnostik diperoleh 78,81%. Persentase siswa yang lulus tes sebesar 70,3% sedangkan persentase siswa yang tidak lulus sebesar 30%. Hal tersebut menunjukkan bahwa instrumen tes diagnostik sudah efektif dimana dapat meningkatkan pemahaman siswa terhadap materi kimia asam basa.

Kata kunci: Pengembangan; Instrumen Tes Diagnostik; Kognitif Belajar; Tes Penempatan; Asam Basa.

INTRODUCTION

Education is a basic need that must be met by every individual throughout life. The curriculum is an educational plan designed to shape the younger generation so that they can play a role and become useful members of society after undergoing formal education at school.

The Independent Curriculum is a curriculum innovation that emphasizes freedom, creativity, and students' adaptability (Gumilar, 2023). This curriculum aims to form independent, resilient, and innovative students through an open and inclusive approach. Therefore, the Independent Curriculum is very important for the success of education and must be developed by the government to keep up with the times. Learning is a combination of a structured series of humans, materials, facilities, tools, and methods that interact with each other to achieve learning goals.

The purpose of learning is to create a teaching and learning process that supports students in achieving optimal learning outcomes (Hamalik, 2010). Effective learning requires students to develop according to their potential. However, not all students can experience maximum or good progress in the learning process. The Independent Curriculum assesses students' cognitive abilities to determine their knowledge abilities during the learning process.

According to cognitive theory, learning is a process involving active mental activity, where students not only store information but students can also process and change it (Gage & Berliner, 1984). Cognitive abilities can improve students' overall qualifications (Vidayanti, 2017). There is a theory that writes about the main ability to think, namely Benjamin Samuel Bloom's theory, where in high-quality education can be achieved and apply all levels of cognition in the learning process (Huda, 2013). The cognitive domain includes behavior that emphasizes all intellectual aspects, namely knowledge and thinking skills, from basic thinking skills such as remembering to high-level thinking skills such as analyzing and drawing.

The United States Office of Education (USOE), namely stating that learning difficulties are disorders experienced by individuals in the early psychological process of understanding and applying lessons or writing. To overcome this problem, a diagnostic approach is used to determine the location of students' learning difficulties. Diagnosis of student misunderstandings can be done by testing. Diagnostic tests are designed to address students' weaknesses and strengths in learning (Ebel, 1979). The main purpose of diagnostic tests is to identify problems faced by students and plan appropriate problem-solving efforts.

Difficulties in understanding chemistry are characterized by complex and abstract concepts and chemicals that cannot be seen. One of the chemistry lessons that is difficult to understand is the material on acids and bases. Acids and bases are materials found in various aspects of everyday life. This material is important because it is closely related to everyday life because its properties can be proven through observation and problem-solving processes.

Based on the results of the interview observations conducted, it is known that students think that in chemistry lessons the acid-base material is difficult. This is because the difficulty in acids and bases is in calculating pH where students still take a long time to calculate pH and also students have difficulty during practicums using pH indicators, universal indicators, and natural indicators. So that students experience a lack of understanding of the concept of acids and bases, which is one of the reasons for the low learning outcomes of students. Based on the description above, it is necessary to analyze students' problem-solving abilities using diagnostic test instruments on acid-base material by grouping classes.

Grouping classes for students with different abilities requires a varied learning approach (Wibowo, 2015). This is based on the assumption that high-achieving students learn faster than students with low abilities. Combining students with different abilities can create a gap in understanding, where students

who learn quickly have to wait for their friends to understand the material. The purpose of this grouping is to increase the effectiveness and efficiency of learning and to facilitate guidance according to the abilities of each student.

Diagnostic tests act as important placement tests in the implementation of the sorogan-bandongan model. Student placement determines the group leader and its heterogeneous members. This approach is different from other learning models, where the group leader functions as an assistant and must be completed before classical learning begins. Grouping by class often uses the Islamic boarding school model such as sorogan and bandongan.

The sorogan bandongan model is taken from Islamic boarding schools in Indonesia. This model helps students understand and learn concepts individually, namely in sorogan where students study the material independently, while bandongan establishes group discussions for those who do not understand. Sorogan and bandongan are variants of learning models which are structured process frameworks in teaching and learning to achieve learning goals (Rinaningsih, 2014).

RESEARCH METHOD

This research is a type of research with a Research and Development (R&D) approach. The development model used is the model according to (Sugiyono, 2019) with the following stages: 1) Potential and Problems 2) Data Collection 3) Product Design 4) Product Validation 5) Design Revision 6) Product Trial 6) Product Revision 7) Usage Trial 8) Product Revision 9) Final Product. This research is adjusted to the research needs. This study was conducted in May 2025 at a public high school in Surabaya. The sample in this study was 30 students of class XI of high school.

Data Collection Techinue

The data collection techniques used in this study consisted of a diagnostic test consisting of 20 question items, a questionnaire consisting of an expert validation questionnaire and a student response questionnaire and

interviews. The diagnostic test instrument was used to determine the validity value of the question items, the level of difficulty, the discriminatory power, and the interpretation of the results of the diagnostic test instrument. The expert validation questionnaire was used to determine the feasibility of the questions before being tested on students while the student response questionnaire was used to see students' responses to the level of readability and the time used in working on the questions. Interviews were conducted to find out how teachers responded to chemistry learning and the diagnostic test instrument as a placement test.

Data Analysis Techiques

The initial stage in data analysis is testing the feasibility or validity of the diagnostic test instrument that has been developed. After the diagnostic test instrument is considered feasible by the expert validator and meets the criteria for instrument feasibility, then the test instrument is tested on students to determine the cognitive level of students in the acid-base material. A score of 1 is given if the answer is correct. While a score of 0 is given if the answer is wrong. The data analysis techniques used include validity, level of difficulty, discriminatory power and interpretation of diagnostic test results.

The validity of the test items is to determine whether the test items from the diagnostic test instrument that has been given meet the valid or invalid criteria. The validity of the items for multiple choice questions can use the point biserial formula in. (Silitonga, 2014)

$$\text{R}_{\text{pbis}} = \frac{Mp - Mt}{St} \sqrt{\frac{p}{q}} (1)$$

the validity coefficient obtained (rpbis) is compared with the rtable product moment values at $\propto = 0,05$. If rpbis > rtable then the test item is said to be valid.

The level of difficulty is done to find out the level of difficulty of the questions. The level of difficulty can be known through the formula below with the criteria for the level of difficulty in table 1. (Arikunto, 2012)

$$P = \frac{B}{JS}$$

Table 1. Difficulty Level Criteria

Difficulty Level Index	Interpretation
0.00 – 0.30	Difficult
0.31 – 0.70	Currently
0.71 – 1.00	Easy

(Rusilowati, 2014)

The discriminatory power can be determined using the formula below with the discriminatory power criteria as in table 2. (Arikunto, 2012)

$$D = \frac{Ba}{Ja} - \frac{Bb}{Jb} = Pa - Pb$$

Table 2. Distinguishing Power Criteria

Item discrimination index	Interpretation of the discriminating power of questions
$D \geq 0.40$	Very good
$0.30 \leq D \leq 0.40$	Good
$0.20 \leq D \leq 0.30$	Enough
$D < 0.20$	Bad

The Student Response Questionnaire can be known through the formula below with the response questionnaire criteria as in table 3.

$$P = \frac{S}{N} \, x \, 100\%$$

Table 3. Percentage of Student Response Scoring Criteria

Percentage	Criteria
82% < score ≤ 100%	Very good
64% < score ≤ 82%	Good
49% < score ≤ 64	Not good
% < score ≤ 49	Not good

Interpretation of diagnostic test results

This research was limited to a limited trial phase of the diagnostic test instrument. Before obtaining diagnostic test scores, a single-group pretest and posttest were conducted with a sample of only one class without a comparison, as described below.

$$\boxed{\text{O1 X O2}}$$

Description:

O1: Pretest score

O2: Posttest score

X: Classroom treatment

The effectiveness of the diagnostic test was obtained from the pretest and posttest sheets. After that, the improvement in learning outcomes was determined using the N-Gain formula as follows:

$$N - gain = \frac{Posttest \ Score \ - \ Pretest \ Score}{Ideal \ Score - Pretest \ Score}$$

The N-Gain score obtained is interpreted into the following criteria:

Table 4. N-Gain Score Criteria

N-Gain Value	Information
g > 0,7	High
0,3 ≤ g ≤ 0,7	Medium
g < 0,3	Low

The analysis of the interpretation of the results uses binary assessment, with the interpretation of the results can be seen in following table;

Score	Information
1	Correct
0	Wrong

Determine the degree of understanding by using the percentage of student assessment using a formula as in the equation below.

$$P = \frac{F}{N} \, x \, 100\%$$

The results of this percentage calculation are then grouped into frequency distributions and percentages as in table 6.

Table 6. Criteria for student learning outcome scores

No	Mark	Category
1.	85-100	Very high
2.	65-84	Tall
3,	55-64	Currently
4.	35-54	Low
5.	0-34	Very low

RESULT AND DISCUSSION

The development of this test instrument is based on the need for a measuring instrument that not only evaluates the final results of learning, but is also able to describe the initial conditions and thinking processes of students. In chemistry learning, especially regarding acid-base material, students often face difficulties in understanding abstract concepts such as acidity levels, pH, ionization reactions, and acid-base theories (*Arrhenius, Bronsted-Lowry, and Lewis*), therefore, the existence of this diagnostic test instrument is very important to help teachers place students according to their level of cognitive ability, as well as a foundation in formulating appropriate learning strategies.

Validation

Experts in the field validated the diagnostic test instrument that was designed to produce good content validity questions. The results of the validation of the diagnostic test instrument showed that there were a number of questions that had to be improved first so that they were in accordance with the indicators that were to be achieved. Product improvements were made by looking at input or suggestions. validator and those who have theoretically valid various aspects of material, construction aspects, language aspects, and other aspects that help the appearance of the product.

Validation data obtained by the validator showed that 25 questions were considered valid . Of these, 14 questions were included in the very good criteria, meaning they were used without improvement with a score between 14 and 19 questions. 10 questions were included in the good category, meaning they still needed a little improvement with a score between 9 and 14.

Furthermore, the diagnostic test instrument that had been validated would go through small-scale and large-scale trials to detect students' knowledge levels. The trial use was carried out in class XI of high school, with a working time of 90 minutes.

The validity test of the test items is used to determine the validity of the questions in the diagnostic test instrument. The test items are said to meet the valid criteria if r count > r table. Based on the validity test conducted on 30 students using SPSS version 16.0 and the r table value of 0.3809, all 25 test items were declared valid.

The difficulty level test was conducted to identify the level of difficulty of the questions. Whether the questions are easy, medium , or difficult. Based on the data from the test results of the level of difficulty of the questions, it can be concluded that every 11 questions are included in the easy, criteria while 14 items are included in the medium criteria.

The discriminatory power test aims to assess the ability of questions to differentiate between students with high and low abilities. Based on the analysis using SPSS version 16.0, all 25 questions have very good discriminatory power so they are suitable for use.

Table 7. Summary of Trial Results

Test Results	Information
Item Validity Test	25 valid questions
Difficulty Level Test	11 questions include easy criteria
	14 questions include medium criteria
Discriminatory Power Test	25 The test items are included in the very good criteria

Practicality

The response questionnaire aims to measure students' responses to the diagnostic test instruments provided. The questionnaire contains questions about the readability of the questions

and responses to the instrument as a placement tool in measuring the cognitive level of student learning.

Table 8. Student Response Questionnaire Results

No	Assessment aspect	percentage (%)	Category
1.	Readability of test questions	76	Good
2.	Ease of understanding of test questions	71	Pretty good
3.	Appropriate length of sentences in test questions	70	Pretty good
4.	Readability of sentences in test questions	75	Pretty good
5.	Ease of understanding of test questions	71	Pretty good
6.	Freedom of questions in test questions to allow for multiple interpretations	70	Pretty good
7.	Readability of images or tables in questions	75	Pretty good
8.	Ease of understanding of images or tables in questions	74,16	Pretty good
9.	Appropriate number of questions given	75	Pretty good
10.	Appropriate time allotted for completing test questions	77,5	Good
average amount		73,42%	

Based on the results of the student response questionnaire assessment, the average percentage of questionnaire results reached 73.42%, which is classified as good.

Effectiveness

The effectiveness of the diagnostic test was obtained from the pretest and posttest sheets. The pretest and posttest consisted of 16 multiple-choice questions. The pretest and posttest aimed to determine improvements in student learning outcomes. Pretest and posttest scores can be calculated using N-gain. Improvements in student learning outcomes were analyzed using the N-Gain score. Table 9 below shows the results of calculations using the N-Gain score on the student knowledge test.

Table 9. Knowledge Test Results for Each Individual with Student N-Gain Score

Students	Pre test	Post test	N-gain Score	Category
1	31	100	1	High
2	75	100	1	High
3	38	100	1	High
4	75	100	1	High
5	31	100	1	High
6	81	100	1	High
7	94	100	1	High
8	63	100	1	High
9	25	100	1	High
10	6	100	1	High
11	88	100	1	High
12	50	100	1	High
13	63	100	1	High
14	69	81	0,38	Medium
15	44	100	1	High
16	25	63	0,50	Medium
17	50	94	0,88	High
18	100	100	0	no increase occurred
19	50	100	1	High
20	25	100	1	High
21	25	100	1	High
22	81	100	1	High
23	88	100	1	High
N-Gain Learning Outcomes			1	High

In table 9, the N-Gain score obtained is 1, which can be said to have increased by 100% with high criteria.

Figure 1 shows the N-Gain results of 23 students who obtained N-Gain with high criteria of 20 students or 86.95%, medium criteria of 2 students or 8.7% and criteria of no increase of 1 student or 4.34%.
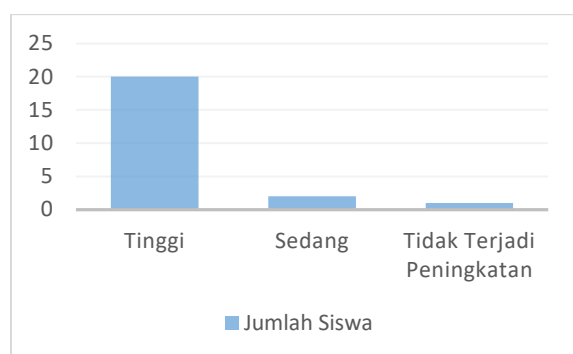

Figure 1. Graph of N-Gain Student Learning Outcomes

Student learning outcomes improved after reading the acid-base handout using the Sorogan learning method, as evidenced by their posttest and N-Gain scores. The average N-Gain score fell within the high criteria.

After completing the pretest and posttest, students completed a diagnostic test. The diagnostic test consisted of 25 multiple-choice questions. The diagnostic test scores are shown in Table 10 below.

Table 10. Diagnostic Test Instrument Results

Number of Studnets	Score	Information
1	24	Not Pass
2	76	Pass
3	92	Pass
4	12	Not Pass
5	76	Pass
6	76	Pass
7	64	Not Pass
8	0	Not Test
9	88	Pass
10	0	Not Test
11	100	Pass
12	72	Not Pass
13	88	Pass
14	96	Pass
15	56	Not Pass
16	76	Pass
17	88	Pass
18	84	Pass
19	72	Not Pass
20	72	Not Pass
21	80	Pass
22	0	Not Test
23	72	Not Pass
24	96	Pass
25	96	Pass
26	96	Pass
27	80	Pass
28	80	Pass
29	100	Pass
30	96	Pass

The results of the interpretation of the diagnostic test instrument stated that out of a total of 30 students as research subjects, 27 students took and completed the test, while 3 students did not take the test. The analysis showed that 19 students obtained a percentage of 70.3% successfully getting a score above the KKM, which was 75 and were declared to have passed or completed. Meanwhile, 8 students obtained 30% still below the limit or completion or below the KKM and were declared not yet completed, as can be seen in table 11.

Table 11. Percentage of completion of chemistry learning outcomes of grade XI high school students

No	Category	Frequency	Percentage (%)
1.	Not finished	8	30
2.	Completed	19	70.3
Amount		27	100.3

The average value of the chemistry test results of students who have used the diagnostic test instrument via the quiziz web is 78.81%, which shows that in general the cognitive level of students is at a high level in understanding acid-base material. The majority of students are

in the middle to upper range, which indicates that most have mastered the basic competencies of the material, but there are still groups of students who need more attention to achieve completeness.

Based on the diagnostic test results completed by students, heterogeneous grouping was implemented, taking into account the distribution of scores across the low, medium, and high categories. The goal of this strategy is to create an inclusive learning group dynamic, where high-achieving students can mentor less-able or low-scoring students, and average-ability students can deepen their understanding through discussion and team collaboration.

Classification of students into six groups, which are arranged by combining students based on high and low achievement scores, can be presented in table 13 below.

Table 12. Student Grouping

Student Code	Group 1	Student Code	Group 2	Student Code	Group 3
KS 11	100	KS 29	100	KS 30	96
KS 4	12	KS 1	24	KS 15	56
KS 14	96	KS 3	92	KS 17	88
KS 20	72	KS 23	72	KS 2	76
KS 21	80	KS 27	80	KS 28	80
Jumlah	360		368		396
Rata-rata	72		73,6		79,2

Student Code	Group 4	Student Code	Group 5	Student Code	Group 6
KS 26	96	KS 25	96	KS 24	96
KS 7	64	KS 12	72	KS 19	72
KS 13	88	KS 9	88	KS 18	84
KS 5	76	KS 6	76	KS 16	76
	324		332		328
	81		83		82

The average scores of the six groups formed showed significant differences. Group 1 had an average of 72, group 2 as much 73.6, group 3 as much 79.2, group 4 as much 81, group 5 as much 83, and group 6 as much 82. The differences in average scores between these groups indicate that each group consists of members with varying academic abilities. There was no similarity in average scores reflecting equality between the groups. Conversely, the increasing trend in average scores from group 1 to group 6 indicates that the grouping was based on gradual levels of ability. Thus, it can be concluded that the groups formed were heterogeneous in the context of intergroups.

CONCLUSION

The conclusion of the research on the development of diagnostic test instruments as placement tests to determine the level of cognitive learning of students in acid-base material is explained as follows:

1. The validity of the developed test instrument 25 questions are included in the good category from the validation results by instrument experts. The validity of the diagnostic test questions shows that 25 questions are valid. The level of difficulty results in every 11 questions being included in the easy category and 14 questions being included in the moderate category. The response questionnaire shows that the percentage has 73.42% with a good category.

2. Based on the results of the discriminatory power for diagnostic test questions as placement tests, there are 25 questions that have very good discriminatory power. So

this diagnostic test instrument is suitable for use.

REFERENCES

Arikunto, S. 2012. *Prosedur Penelitian Suatu Pendekatan Praktik.* Jakarta: Rineka Cipta.

Ebel, R. L. 1979. *Essentials of Educational Measurement.* New Jersey: Pratince.

Gage, N. L., & Berliner, D. 1984. *Educational Psychology.* Boston: Hiughton Mifflin Company.

Gumilar, G. 2023. *Urgensi Penggantian Kurikulum 2013 Menjadi Kurikulum Merdeka.* Jurnal Papeda*, 5(2), 148-155.

Hamalik, O. 2010. *Proses Belajar Mengajar.* Jakarta: PT. Bumi Aksara.

Huda, M. 2013. *Model - Model Pengajaran dan Pembelajaran.* Yogyakarta: Pustaka Pelajar.

Rinaningsih, R. 2014. *Implementasi Model Perkuliahan Terpadu Sorogan Bandongan Untuk Menentukan Pemahaman Mahasiswa Dalam Mempelajari Mekanisme Reaksi.* Jurnal Pengajaran MIPA, 19(2), 266-274.

Rusilowati, A. 2014. *Pengembangan Instrumen Penilaian.* Semarang: UNNES PRESS.

Sugiyono. 2019. *Metode Penelitian Kuantitaf, Kualitatif dan R&D.* Bandung: ALFABETA.

Silitonga, P.M., 2014. *Statistika Teori Dan Aplikasi Dalam Penelitian Edisi Kedua,* Fakultas Matematika dan Ilmu Pengetahuan Alam. UNIMED, Medan.

Vidayanti, N. 2017. *Analisis Kemampuan Kognitif Siswa Kelas VIII SMP Negeri 11 Jember Ditinjau Dari Gaya Belajar Dalam Menyelesaikan Soal Pokok Bahasan Lingkaran.* Jurnal Matematika dan Pendidikan Matematika, 8(1), 137-144.

Wibowo, H. 2015. *Penerapan Pengelompokkan Siswa Berdasarkan Prestasi di Jenjang Sekolah Dasar.* Jurnal Psikologi Universitas Diponegoro, 14(2), 148-159.