Journal of Turkish Science Education, 2025, *22*(1), 48-62.
DOI no: 10.36681/tused.2025.004

The development of an integrated instrument to measure higher order thinking skills and scientific attitudes

Betzy Ayu Omega Rampean[1], Eli Rohaeti[2]

[1]*Universitas Kristen Indonesia Paulus, Indonesia, Corresponding author, betzyayu@gmail.com, ORCID ID: 0000-0002-0676-2618*
[2]*Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia, ORCID ID: 0000-0002-0930-732X*

ABSTRACT
Assessment in education involves collecting and processing information related to student achievement during learning. Assessment activities help educators to find out where learners are having difficulties which for the sake of completeness requires an integrated instrument. The assessment carried out should conform to 21st-century learning criteria pertaining to cognitive and affective aspects of learning, and involve the identification of difficulties related to higher-order thinking skills (HOTS) and scientific attitudes of students. The purpose of this research was to develop an integrated assessment instrument that could be used to measure students' HOTS and scientific attitudes in relation to a chemistry topic, reaction rates. The research uses a 4-D development model, which includes the stages of define, design, develop, and disseminate. The results of the research were analysed quantitatively and qualitatively using item response theory (IRT). Based on the content validity, the integrated instrument was declared to be valid according to Aiken's V equation. All items in the test instrument were valid based on the Rasch, item fit, and item difficulty. The integrated instrument has also been applied based on the reliability of the estimated test and test information function (TIF) curves used to determine the profile of students' HOTS and scientific attitude.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
01.04.2021
Accepted:
02.07.2024
Available Online:
19.02.2025

KEYWORDS:
Integrated instrument, HOTS, scientific attitude, chemistry.

To cite this article: Rampean, B. A. O. & Rohaeti, E. (2025). The development of an integrated instrument to measure higher order thinking skills and scientific attitudes. *Journal of Turkish Science Education, 22*(1), 48-62. DOI no: 10.36681/tused.2025.004

Introduction

Learning is a process of developing the potential and character building of each learner as a result of the synergy alls of education that takes place in schools, families and communities in the hope that students can solve various problem that occur around them (Yunus & Ali, 2013). Andreas Schleicher, Director of Education and Skills of the Organisation for Economic Co-operation and Development (OECD), said that this is the time for Indonesia to raise the quality of teaching and learning to ensure that people have various skils (Minister of Education and Culture Regulation). Superior human resources (HR) were needed in facing the progress of the 21st century The National Education Association has identified 21st-century skills as "The 4Cs", which includes critical thinking, creativity, communication and collaboration abilities (Changwong et al., 2018). In fact, pupils in senior

high schools in Indonesia tend to exhibit average scores in the aspect of communicating because they are reluctant to ask questions during class discussions (Sukardiyono et al., 2019).

Pre-service teachers in Indonesia tend to score poorly on critical thinking skills (Irwanto et al., 2019). Students are not only asked to apply what they have learned but also to analyse and evaluate what they have acquired to be able to solve problems and make decisions in everyday life (Pratama & Retnawati, 2018). To sharpen these skills, pupils can build the learning process by implementing HOTS-based evaluation questions (Yulianto et al., 2019). Students need a long time to complete HOTS tests (Rampean, et al., 2022). The problems must be identified first, the reason for doing the work, the hypotheses or theoretical predictions under consideration, and the essential background. It should not contain equations or mathematical notation.

Student's scientific attitudes can be developed during the learning process in the classroom such as curiosity, honesty, cooperation, open mind and ideas, diligence and conscientiousness (Mayang et al., 2020). These activities include part of a scientific attitude that is open to thoughts and ideas, building their knowledge by working together between teachers and learners as well as between learners and learners to exchange experiences during classroom learning activities (Anggraeni et al., 2017). Budiharti and Waras (2018) also state that scientific attitude has its own dimensions, being curiosity, respect for data/facts, critical thinking, discovery and creativity, open-mindedness and cooperation, persistence, and sensitivity to environmental changes. A scientific attitude is an attribute that scientists must have to do their jobs A scientific attitude is needed by learners to see when theory meets practice (Ataha & Ogumogu, 2013). so that they can re-explore their initial knowledge and interpret what they have experienced.

Learning success is measured by assessment activities. Assessment needs to be comprehensive (Atkin & Coffey, 2003; Darmawan et al., 2020). Diagnostic assessment is closely related to evaluating learners' ability to achieve learning goals. Assessment instruments such as test instruments must be developed with proper planning as a strategic step to solve the problems (Adams et al., 2015; Gurcay & Gulbas, 2015; Kirschner et al., 2016).

The assessment carried out by educators must be in accordance with the educational assessment standards stipulated by the Regulation of the Minister of Education and Culture Number 23 of 2016 that "arrangements regarding educational assessment need to be adjusted to developments and needs in the assessment of learning outcomes". Thus educational assessment must be adjusted to the development and needs of students today, namely 21st-century learning which leads to the measurement of higher-order thinking skills and scientific attitudes.

Lai and Hwang (2014) explained that HOTS have been used for question formulation by various government organisations related to education in Singapore, Taiwan and the United States. Indonesia needs to catch up in order to improve its performance on international tests, especially the Programme for International Student Assessment (PISA) (National Education Standards Agency, 2018).

Ichsan et al. (2019) stated that the results of measuring high order thinking skills from primary school children to postgraduate students remain comparatively low. The same thing was also expressed by Gendenjamts (2023) who noted that performance on the creating skill tasks was lower than on the analysing and evaluating tasks. It has been shown that in osmoregulation, PjB spell out-HOTS based learning can develop higher level thinking skills such as creative thinking skills, analytical thinking skills and metacognitive thinking skills (Liline, et al., 2024).Koomson et al. (2024) explained science process skills such as problem solving, critical thinking and decision-making, among others, are related to SPS spell out, including observing, inferring, predicting, classifying, evaluating, experimenting and others classified as SPS in science education. The observation that these important skills are inadequately acquired and developed among teacher trainees, who are required to be future facilitators of these skills among basic school learners, needs immediate attention.

The objectives: to develop an integrated instrument that will measure higher-order thinking skills and scientific attitudes of students using reaction rate as the topic and to determine the content validity, empirical validity, reliability of the integrated instrument.

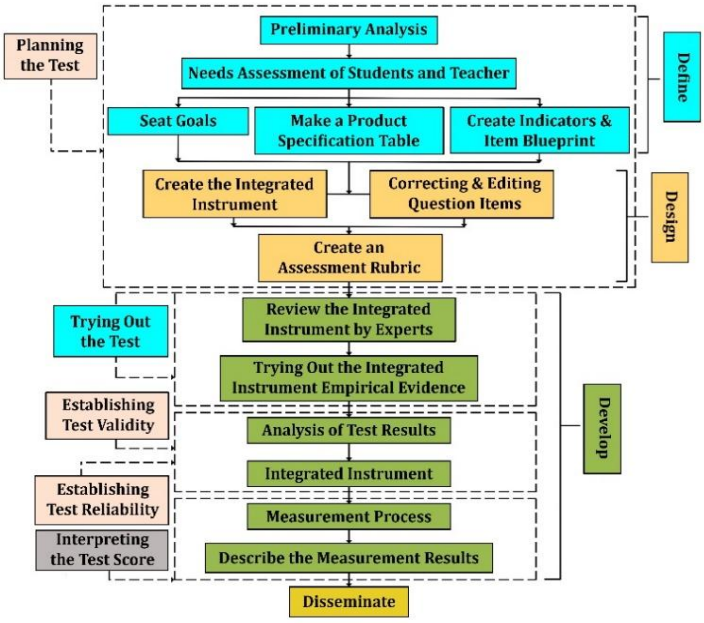
Methods

This study comes under the umbrella of research and development (R & D), which aims to produce a product, in this instance in the form of a set of instruments in order to evaluate students' higher order thinking skills. The research method used is a combination of the 4-D model and the Oriondo & Dallo-Antonio (1984) development model which includes planning the test, trying it out, establishing test validity, establishing test reliability, and interpreting test score.

The main product of this research is an integrated test instrument that can be used to diagnose higher order thinking skills and scientific attitudes in a high school chemistry topic, namely reaction rate. Data analysis techniques include using qualitative and quantitative descriptive analysis. The qualitative analysis aims to see the construction of instruments through expert judgment this is part of the procedure to devise a test, whereas 'analysis' by implication involves the outcomes of that test. Quantitative analysis is used to determine the validity does this not also involve value judgements? and reliability of the instrument. The draft instrument was created and developed into questions, and the was given to experts for content validity that's right, not 'quantitative analysis'. The test instrument consisted of 15 essay questions with cognitive levels at levels C4 (analyse), C5 (evaluate), and C6 (create) note that these are from Bloom's Revised Taxonomy 2001 for the measurement stage the whole test 'measures', 8 items have been selected which represent each learning indicator explain.

Figure 1

Integrated instrument development procedure



Participants

The test instrument was piloted in September 2020 involving 238 pupils of class XI MIPA (Mathematics and Natural Sciences) from three high schools located in Yogyakarta City. The three schools are SMA Negeri 3 Yogyakarta consisting of two classes, SMA Negeri 4 Yogyakarta of three classes, and MA Negeri 1 Yogyakarta consisting of two classes. These schools were chosen to represent each school level, namely low, medium and high based on national test scores.

Data Collection

Instruments that have been developed as indicators of higher order thinking skill and scientific attitude (Appendix 1) were validated (content validity) through expert judgment using the instrument item review criteria (Appendix 2) (Pujayanto et al., 2018). The assessment was conducted by seven raters, namely assessment experts, two chemistry learning experts and four practitioners. The instrument readability test was carried out by giving the instrument to 238 students. Then, the validity for empirical evidence was gathered.

Data Analysis

The content validity of the test instruments was established using Aiken's V formula. The coefficient of content validity was based on the judgment of the experts. Aiken's V index value is calculated as

$$V = \frac{\sum s}{[n(c-1)} = \frac{\sum(r\text{-}lo)}{[n(c-1)]} \quad (1)$$

n is the number of raters, s is the value from experts judgment, c is the highest validity rating, lo is the lowest validity rating, r is the number given by the rater (Aiken, 1985; Azwar, 2012). Azwar (2015) explained that the validation criteria value of the integrated test instrument for higher order thinking skill and scientific attitude is divided into five categories as shown in Table 1.

Table 1

Criteria value of Aiken's V validity

Validity Value	Category
$0.8 \leq V \leq 1.0$	Very Good
$0.6 \leq V \leq 0.8$	Good
$0.4 \leq V \leq 0.6$	Acceptable
$0.2 \leq V \leq 0.4$	Bad
$V \leq 0.2$	Very Bad

Validity test show empirical evidence is obtained through analysis of item responses to test results in the form of polytomous data. Polytomous data were analysed using Item Response Theory (IRT) according to the Rasch model or Partial Credit Model (PCM) 1 Parameter Logistics (1-PL). Analysis using the Quest and Excel programmes. The Quest programme is used to determine the goodness of fit, reliability, and item difficulty index. The Excel programme is used to designate information of the variance-covariance matrix between groups of students' abilities. Linacre (2012) explained that items that fit or were categorised fit with the PCM model if the mean and standard deviation of INFIT MNSQ were between 0.5 until 1.5 and the INFIT t value was between -2.0 until 2.0 (Hambleton & Swaminathan, 1985).

The reliability of the test estimate on the integrated instrument can be determined using the Quest programme. The reliability results are known from the output data with the extension sh (.sh) in the Summary of Item Estimates section. Subali (2011) states that the higher the reliability coefficient the more reliable the instrument and the smaller the possibility of errors. George and Mallery (2020) categorised the reliability coefficients as shown in Table 2.

Table 2

Interpretation of Reliability

Reliability Coefficient	Category
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$V < 0.5$	Unacceptable

The product user response was in the form of instrument readability by students obtained through a questionnaire that had been validated by expert judgment. The final score was converted into properness categories using the guidelines shown in Table 3 (Sumadi et al., 2015).

Table 3

Criteria value of readability

No	Score Range	Category
1	Xi + 1.8 Sbi < $\overline{X}$	Very Good
2	Xi + 0.6 Sbi < $\overline{X}$ ≤ Xi + 1.8 Sbi	Good
3	Xi – 0.6 Sbi < $\overline{X}$ ≤ Xi + 0.6 Sbi	Acceptable
4	Xi – 1.8 Sbi < $\overline{X}$ ≤ Xi – 0.6 Sbi	Bad
5	$\overline{X}$ ≤ Xi – 1.8 Sbi	Very Bad

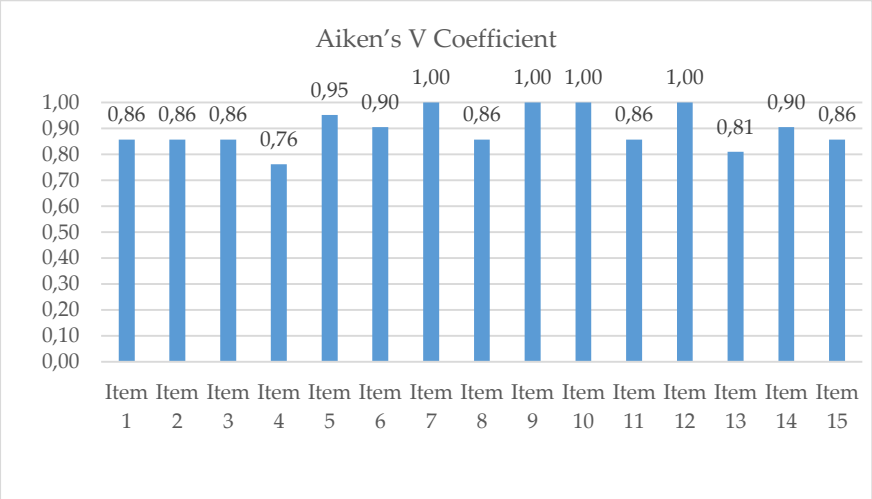
Findings

Content Validity

The content validity of the developed test instruments was reviewed by seven raters with four rating scales before administering the instrument. The results of the analysis are shown in Figure 2.

Figure 2

Argumentative skills score of PSTs



Based on the standard set by Aiken, the minimum standard of Aiken's V coefficient for this study is 0.76 with a probability of 0.045 (Aiken, 1985).

Empirical Validity

The results of the goodness of fit analysis can be seen from the INFIT parameters for Mean Square (MNSQ) showing that the integrated instrument for diagnosing higher order thinking skills and scientific attitude meets the statistical fit criteria presented in Table 4.

Table 4

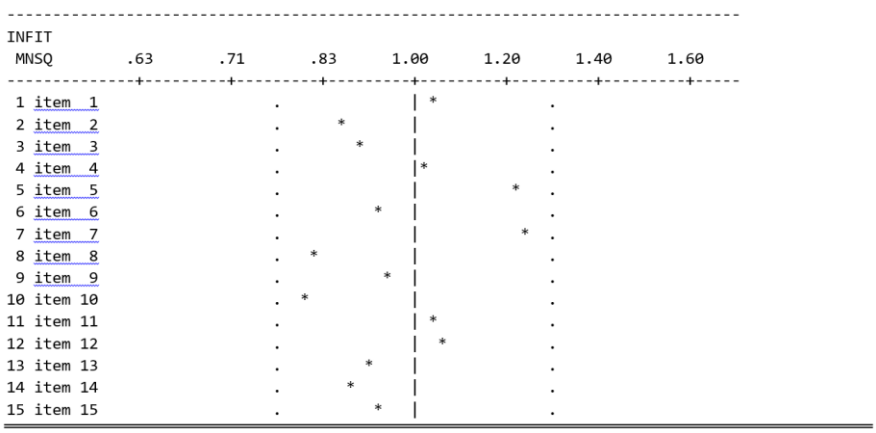
Fit model at 0.5 probability level

No. Item	INFIT MNSQ	Interpretation
Item 1	1.04	Fit
Item 2	0.87	Fit
Item 3	0.89	Fit
Item 4	1.01	Fit
Item 5	1.23	Fit
Item 6	0.93	Fit
Item 7	1.23	Fit
Item 8	0.82	Fit
Item 9	0.94	Fit
Item 10	0.81	Fit
Item 11	1.05	Fit
Item 12	1.06	Fit
Item 13	0.91	Fit
Item 14	0.88	Fit
Item 15	0.93	Fit

The analysis result shows that the item's estimated reliability is 0.81-1.23. which means the test sample suitable the item tested and is very good as referred to in Table 4. or the sample provides consistent results and information as expected. The suitability map of 15 items with the Rasch model is shown in Figure 3.

Figure 3

Map fit model



An item is said to be "good" if it has a good level of difficulty, namely -2 logit ≤ bi ≤ +2 logit (Hambleton & Swaminathan, 1985). The results of the analysis show that 15 items fall into the good category.

Difficulty Level

A good question is a question that has met the requirements of validity and reliability, and has a balance of question difficulty levels. The following are the results of the analysis of question difficulty that has been carried out in table 5.

Table 5

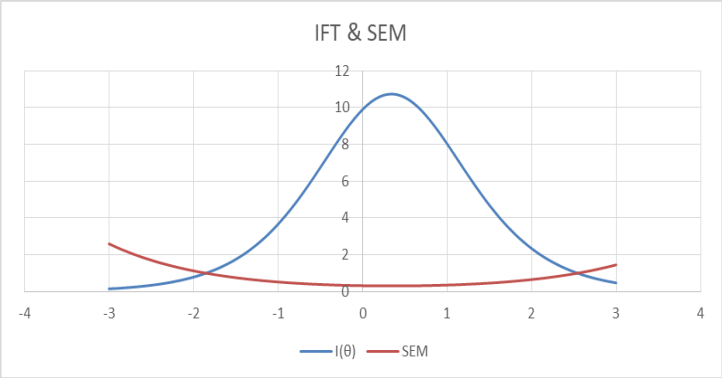
Difficulty level

No. Item	Item difficulty	Category
Item 1	0.670	Moderate
Item 2	0.390	Moderate
Item 3	0.525	Moderate
Item 4	0.323	Moderate
Item 5	0.353	Moderate
Item 6	0.254	Difficult
Item 7	0.317	Moderate
Item 8	0.254	Difficult
Item 9	0.293	Difficult
Item 10	0.287	Difficult
Item 11	0.296	Difficult
Item 12	0.283	Difficult
Item 13	0.266	Difficult
Item 14	0.267	Difficult
Item 15	0.393	Moderate

Based on table 5, it is known that as many as 7 questions or as many as 47% of questions fall into the moderate category, 8 questions or as many as 53% of questions fall into the difficult category, and there are no questions that fall into the easy category.

Figure 4

Relationship between TIF and SEM



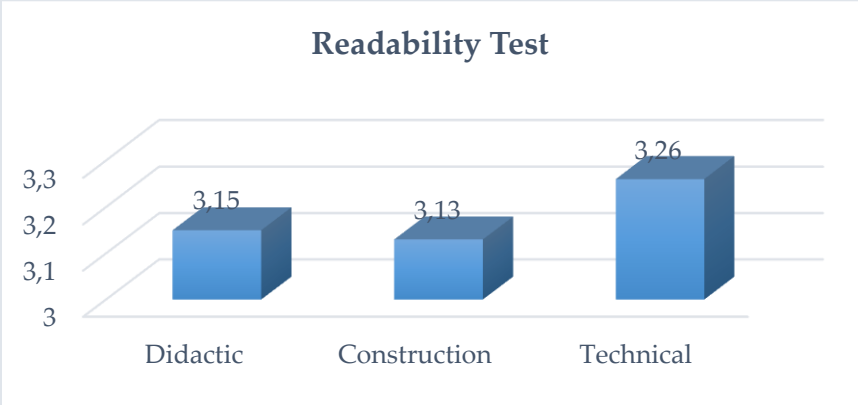
Based on the results of the analysis using the Rasch model used has a maximum test information function value of 10.74362 at θ around +0.35 logit.

Readability Test

The readability test is to measure the text level. It will be able to predict comprehension at the sentence level. Readability as the degree to which what is represented by written text is simple or difficult to comprehend shown in Figure 5.

Figure 5

Readability test



Results instrument readability test (in terms of didactic, construction and technical aspects and suitability of questions presented with chemistry material) involving 238 students showed good results without the need to be revised. Thus, this test instrument is suitable for use

Discussion

The empirical data was conducted to find out the indicators used in higher order thinking skill and scientific attitude. The development of this test instrument used chemical reaction rate as a topic divided into sub-topics namely the concept of rate of reaction, equation of reaction rate and the factors that influence reaction rate. This research has synthesised indicators of higher order thinking skill and scientific attitude of several experts and then integrating them into new indicators.

The content validity of the developed test instruments was reviewed by seven raters with four rating scales before testing the instrument. Based on the standard set by Aiken, the minimum standard of Aiken's V coefficient for this study is 0.76 with a probability of 0.045 (Aiken, 1985). The content validity was analyzed quantitatively and the results of the analysis are shown in Figure 2. Figure 2 shows all items between 0.81 to 1.00 and has exceeded the minimum of Aiken's V coefficient limits. All instrument items can be declared valid based on content validation analysis using the Aiken's V coefficient and all items are very good as referred to in Table 1.

Table 5 shows the 15 item questions have a difficulty level in the range 0.254 - 0.670 in the good category. Retrieval of data using google form so that it allows students to cheat when working on question items also greatly affects the difficulty level of the item questions and the level of difficulty that has been formulated by the teacher is not in accordance with the difficulty level of the empirical results (Amelia & Kriswantoro, 2017). This is because in making these items, the teacher classifies items into a certain level of difficulty (easy, medium, and difficult) based solely on their intuition. The quality of the test item or not can be seen from the difficulty level of each item, the test item is said to be good if the item is neither too difficult nor too easy, or in other words, the difficulty

level of the item is moderate or sufficient. Hambleton (in Wiberg, 2004) stated a good (reliable) test has a TIF value of ≥ 10. If seen in table 2, the reliability of the test is included in the excellent category, so this instrument is reliable for measuring students' chemical abilities. Meanwhile, SEM is inversely related to the test information function. This means that the test will provide good information, with the smallest measurement error being 0.305.

Figure 5 in the didactic aspect the average obtained is 3.15, which is included in the good category. This reflects that the didactic aspect of the integrated instrument that has been developed requires high-level thinking skills and the scientific attitude of students to answer questions, the questions developed are also in accordance with daily life, easy to understand questions and questions that have not been or rarely encountered before. The average construction aspect ??? is 3.13, which is in the good category. These results reflect the language aspects of the integrated instrument that have been developed that do not cause multiple interpretations, provide sufficient space to write down each answer, write item questions in the instrument using enhanced spelling, the language used is simple, straightforward and easy to understand. The average on the technical aspect is 3.26, which is included in the good category. These results reflect the display aspects of the integrated instrument developed that have proportional types and sizes of letters used, chemical symbols written correctly and are easy to understand, the presentation of images and tables in the instrument is able to help students remember questions that have been read, all pictures in the instrument is able to clarify each core question, the combination of colours, images, and writing on the instrument is compatible and according to the proportional size.

Conclusion and Implications

The integrated assessment measure the higher order thinking skills and scientific attitude of senior high school students instrument has met Aiken's validity content based on expert judgment with valid criteria. Students responses to the integrated assessment instrument were stated to be good. The empirical validity phase based on the lowest and highest limits of MNSQ 0.5 and 1.5, all test items (15 items) were declared fit with PCM 1-PL. Test reliability was 10.74 in θ same +0.35 logit with the excellent category. The difficulty level of question item are difficult and moderate, as they range from 0.254 to 0.670. Therefore, the integrated assessment instrument is suitable for measuring the higher order thinking skills and scientific attitude of senior high school students.

This study was conducted on final year students in three schools in Yogyakarta on the material reaction rate and the results stated that in an integrated manner the high-order thinking skills and scientific attitudes of students were in the average category. Conducting this evaluation an integrated assessment instrument to measure the higher-order thinking skills and scientific attitudes of students on other chemical materials by analysing competency standards and basic competencies to match the characteristics of the instrument to be developed.

References

Adams, W. K., Wieman, C. E., Adams, W. K. (2015). Analyzing the many skills involved in solving complex physics problems analyzing the many skills involved in solving complex physics problems. *American Journal of Physics*, 83(459), 1-15. DOI: 10.1119/1.4913923

Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings, *Educational and Psychological Measurument*, 45(1), 131-142. doi: http://doi.org/10.1177/0013164485451012.

Amelia, R. N., & Kriswantoro, K. (2017). Implementation of item response theory for analysis of test items quality and students' ability in chemistry. *Jurnal Kimia dan Pendidikan Kimia*, (2)1, 1-12. ISSN 2503-4146.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman, 268-269.

Anggraeni, T., Sugiyo., & Kustiono. (2017). The difference of ability to ask, scientific attitude. motivation before and after following contextual teaching and learning model. *Journal of Primary Education*, *6*(3), 248-256.

Ataha, U. C., & Ogumogu, A. E. (2013). An investigation of the scientific attitude among science students in senior secondary schools in edo south senatorial district, edo state. *Journal of Education and Practice*, *4*(11), 12-16.

Atkin, J. M., & Coffey, J. E. (2003). Everyday Assessment in The Science Classroom (Issue c). National Science Teachers Association.

Azwar, S. (2012*). Penyusunan Skala Psikologi (Psychological Scale Compilation)*. Pustaka Pelajar.

Azwar, S. (2015). *Reliability and Validity (fourth)*. Pustaka Pelajar.

Budiharti, R., & Waras, N. S. (2018). Analysis of student's scientific attitude behaviour change effects blended learning supported by I-spring Suite 8 application. *Proceedings of Journal of Physics*, Indonesia,*1022*, 1-10. DOI:10.1088/1742-6596/1022/1/012024

Changwong, K., Sukkamart, A., & Sisan, B. (2018). Critical thinking skill development: analysis of a new learning management model for thai high schools. *Journal of International Studies, 11*(2), 37-48. DOI: 10.14254/2071-8330.2018/11-2/3

Darmawan, D., Yatimah, D., Sasmita, K., & Syah, R. (2020). Analysis of Non-Formal Education Tutor Capabilities in Exploring Assessment for Science Learning. *Jurnal Pendidikan IPA Indonesia*, *9*(2), 267–275. DOI:10.15294/jpii.v9i2.24025

Gendenjamts, S. (2023). Measurıng Hıgher Order Thınkıng Skılls in Scıence Among Prımary School Students Usıng Item Response Theory. *European Journal of Education Studies*, *10*(12), 19-28. https://doi.org/10.46827/ejes.v10i12.5089

George, D., & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference (Sixteenth).* Routledge. https://doi.org/10.4324/9780429056765

Gurcay, D., & Gulbas, E. (2015). Development of Three-Tier Heat, Temperature and Internal Energy Diagnostic Test. *Research in Science and Technological Education*, *33*(2), 197–217. https://doi.org/10.1080/02635143.2015.1018154

Hambleton, R. K., & Swaminathan, H. (1985). *Items Response Theory: Principles and Application.* Kluwer-Nijjhoff Publish.

Hastuti, P. W., Nurohman, S., & Setianingsih, W. (2018). The development of science worksheet based on inquiry science issues to improve critical thinking and scientific attitude, *Journal of Physics: Conf. Series, 1097*, 1-8. http://doi.org/10.1088/1742-6596/1097/1/012004

Hilalliati, N., Jumadi, Wilujeng, I., & Kuswanto H. (2019). Scientific attitudes mapping of students after using phet assisted group investigation models. *Journal of Physics: Conf. Series, 1233*, 1-8. http://doi.org/doi:10.1088/1742-6596/1233/1/012050

Hugerat & Kortam. (2014). Improving higher order thinking skills among freshmen by teaching science through inquiry. *Eurasia Journal of Mathematics, Science & Technology Education*, *10*(5), 447-454. https://doi.org/10.12973/eurasia.2014.1107a

Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning. *European Journal of Educational Research*, *8*(4), 935-942. https://doi.org/10.12973/eu-jer.8.4.935

Irwanto, Rohaeti, E., Prodjosantoso, A. K. (2019). Analyzing the Relationships between PreService Chemistry Teachers' Science Process Skills and Critical Thinking Skills. *Journal of Turkish Science Education, 16*(3), 299-313.

Kirschner, S., Borowski, A., Fischer, H. E., Gess-Newsome, J., & von Aufschnaiter, C. (2016). Developing and evaluating a paperand-pencil test to assess components of physics teachers' pedagogical content knowledge. *International Journal of Science Education*, *38*(8), 1343–1372. https://doi.org/10.1080/09500693.2016.1190479

Koomson, A., Christopher, Y. W., & Christine, A. Y. (2024). Effect of Science Process Skills and Entry Grades on Academic Scores of Student Teachers. *Journal of Turkish Science Education*, *21*(1), 118-133. https://doi.org/1 0.36681/tused.2024.007

Kusuma, M. D., Rosidin, U., Abdurrahman., & Suyatna, A. (2017). The development of higher order thinking skill instrument assessment in physics study. *Journal of Research & Method in Education*, *7*(1), 26-32. https://doi:10.9790/7388-0701052632

Lai, C. L., & Hwang, G. J. (2014). Effects of mobile learning participation time on high school students' 21st century core competences. *International Conference of Educational Innovation through Technology*, *1*(4), 205-211. https://doi:10.1109/EITT.2014.40

Liline, S., Anensiana, T., Dominggus, R., & Kristin S. (2024). The Effect of The Pjb-HOTS Learning Model on Cognitive Learning, Analytical Thinking Skills, Creative Thinking Skills, and Metacognitive Skills of Biology Education Students. *Journal of Turkish Science Education*, *21*(1), 175-195. https://doi:10.36681/tused.2024.010

Linacre, J.M. (2012). *A user's guide to winsteps-ministeps rasch model computer program (program manual 3.75.0).* Chicago.

Mayang, F., Yusrizal., Mustafa (2020). Application of guided inquiry learning model to improve students' scientific attitudes and learning outcomes. *Proceedings of Journal of Physics*, Aceh, *1460*, 1-6.

Nath, S., Thomas, S. (2012). Enhancing science process skills and scientific attitude and analysing their interactions: an intervention through inquiry learning approach. *International Journal of Scientific Research, 1*, 37-42.

National Education Association. (2012). *Preparing 21st century students for a global society: an educator's guide to the four cs*. National Education Association.

Oriondo, L. L., Dallo-Antonio, E. M. (1984). *Evaluating Educational Outcomes (Test, Measurement, and Evaluation) (5th Editio).* REX Printing Company, Inc, 279-280.

Permendikbud No. 22 Tahun 2016 tentang Standar Proses Pendidikan Dasar dan Menengah (Minister of Education and Culture Regulation No. 22 of 2016 on Process Standards for Primary and Secondary Education), (2016).

Pratama, G. S., Retnawati, H. (2018). Urgency of higher order thinking skills (hots) content analysis in mathematics textbook. *Proceedings of Journal of Physics*, Bristol, 1097, 1-8. https://doi.org/0.1088/1742-6596/1097/1/012147

Pujayanto, Budiharti, R., Radiyono, Y., Rizky, N., Nuraini, A., Putri, H. V., Saputro, D. E., Adhitama, E. (2018). Pengembangan Tes Diagnostik Miskonsepsi Empat Tahap Tentang Kinematika (Developing Four-Tier Misconception Diagnostic Test about Kinematics). *Cakrawala Pendidikan, 37*(2), 237–249. https://doi.org/10.21831/cp.v37i2.16491

Rampean, B. A. O., Rohaeti, E., Utami, W. P. (2022). Teacher Difficulties forDevelop Higher Order Thinking Skills Assessment Instrument onReaction Rate. *Jurnal Pendidikan Kimia Indonesia*, *6*(1), 11-19 https://doi.org/10.23887/jpk.v6i

Subali, B. (2011). Pengukuran kreativitas keterampilan proses sains dalam konteks assessment for learning. *Jurnal Cakrawala Pendidlikan*, *15*(1), 130-144 https://doi.org/10.21831/cp.v1i1.4196

Sukardiyono, Rosana, D., Dwandaru, W. S. B. (2019). Measuring junior high school students' science learning and science process skills through and integrated science instructional assessment. *Journal of Turkish Science Education*, *16*(4), 467-477.

Sumadi, C.D., Mulyani, S., Agustina, W. (2015). Pengembangan media game senyawa hidrokarbon pada pembelajaran kimia di SMA batik 1 Surakarta dan SMA batik 2 Surakarta. *Jurnal Pendidikan Kimia*, *4*(2), 82-88

Supardi, R., Istiyono, E., Setialaksana, W. (2019). Developing scientific attitudes instrument of students in chemistry. *Journal of Physics: Conf. Series*, *1233*, 1-9. http://doi.org/doi:10.1088/1742-6596/1233/1/012025

Verdina., Gani., Sulastri. (2018). Improving students' higher order thinking skills in thermochemistry concept using worksheets based on 2013 curriculum. *Journal of Physics: Conf. Series, 1088*, 1-6 https://doi:10.1088/1742-6596/1088/1/012105

Wiberg, M. (2004). *Classical test theory vs item response theory.* Umea University.

Yonata, Nasrudin. (2018). Laboratory activity worksheet to train high order thinking skill of student on surface chemistry lecture. *Journal of Physics: Conf. Series, 947*, 1-7. https://doi:10.1088/1742-6596/947/1/012027

Yulianto, T., Pramudya, I., Slamet, I. (2019). Effects of the 21st century learning model and problem-based models on higher order thinking skill. *International Journal of Educational Research Review*, *4*(4), 749-755. https://doi.org/10.24331/ijere.629084

Yunus, F.W., Ali, Z. M. (2013). Attitude towards learning chemistry among secondary school students in malaysia. *Journal of Asian Behavioural Studies*, *3*(11), 1-11.

Appendix 1

Synthesis Results for the Development of Test Instruments

HOTS Theory	Scientific Attitude Theory	Operational Verb
HOTS is a thinking skills that requires the ability to analyse (C4), evaluate (C5), and create (C6), as well as the dimensions of knowledge in the form of factual knowledge, conceptual procedures, and metacognitive (Kusuma, et al., 2017).	Students who carry out scientific activities have an increased ability to think, students are able to interpret the results of investigative activities, besides that students can construct their own knowledge so that students become excited, curious, respect data , critical thinking and have an open mind (Hastuti el al., 2018))	HOTS: Analyse, organise, evaluate, check, conclude, correlate, criticise, create and formulate Scientific attitude: Critical thinking
The components of higher order thinking include Analyse (C4), Evaluate (C5), and create (C6). The description of analysing indicators (C4) is observing phenomena, formulating problems, determining hypotheses, defining variables, determining tools and materials, developing testing procedures, analysing and making conclusions; creating indicators (C6), namely designing experimental procedures and hypothesizing (Yonata & Nasrudin, 2018).	Hilalliati et al. (2019) stated that a good scientific attitude includes curiosity, respect for data and facts, critical thinking, open thinking and collaboration, sensitive to the environment and persistence.	HOTS: Analyse, distinguish, formulate problem, evaluate, check, create and attributing Scientific attitude: Critical thinking, thorough and persistence
Verdina, et al. (2018) stated that the three HOTS indicators are analysing, evaluating, and creating. a. In the analysing section (C4), students are stimulated to analyse the system and environment. b. In the evaluating section (C5), there are pictures that stimulate students to evaluate how energy and mass transfer characteristics occur in closed, open and isolated systems. c. In the creating section (C6), there are pictures that stimulate students to be creative in giving reasons for the image that includes outerwear	Supardi, et al. (2019) stated that scientific attitudes include: 1. Honesty: Writing down the results of observations as they are, not seeing the results of other people's observations, not mixing facts with opinions, and accepting the results of observations. 2. Persistence: Do not give up on doing experiments, have a habit of repeating experiments, and doing practicum activities until the end 3. Thorough: Paying attention to empirical facts, Working carefully in practicum, Delaying decision making until sufficient data has been collected, and Working according to instructions 4. Be critical: Seek as much information as possible, pay attention to data even though it is small, and Do not immediately accept conclusions without strong evidence	HOTS: Analyse, distinguish, evaluate, organize and create. Scientific attitude: Honesty, perseverance, accuracy, critical

HOTS Theory	Scientific Attitude Theory	Operational Verb
Higher order thinking skills, namely identifying controls; planning experiments, and drawing conclusions. The development of student HOTS can be done by developing skills to make conclusions from controlled experiments, identifying suitable controls, and planning controlled experiments (Hugerat & Kortam, 2014).	Nath and Thomas (2012) state that scientific attitudes include: 1. Rationality: Commitment to the value of rationality and acceptance of criticism 2. Curiosity: The desire for the completeness of knowledge and Emphasises the questions in the approach to new situations 3. Open mind: Willingness to revise opinions and conclusions and Desire for new things and ideas 4. 4. Superstition Rejection: Rejection of superstitions and false beliefs and Acceptance of Scientific facts	HOTS: Organize, criticize, correlate and attributing Scientific attitude: Honesty, accuracy and critical thinking
Analysis (C4) includes learning to determine the relevant or important part of a message (differentiating), how the message pieces are organized (organized), and the objectives underlying the message (attributing); evaluating is defined as making judgments based on criteria and standards. The criteria most frequently used are quality, effectiveness, efficiency and consistency. (eg, examining and criticizing) and creating involves bringing together elements to form a coherent or functional whole (Anderson & Krathwohl, 2001).	All dimensions of a scientific attitude are curious, respect data or facts, critical thinking, discovery and creativity, open-mindedness and cooperation, and persistence (Budiharti and Waras, 2018).	HOTS: Distinguish, organize, evaluate, check, criticize and create. Scientific attitude: Critical thinking and persistence

Appendix 2

The instrument item review criteria

Aspect		Review Criteria
Material	1	The item matches the indicators of achievement
	2	The item matches the indicator in the item blueprint
	3	The item matches the basic competency
Construction	4	Item formulated with short, dense clearly and unambiguously
	5	Item formulated with short, dense clearly and unambiguously
	6	Homogeneous and logical answer choices.
	7	Homogeneous and logical answer choices.
	8	Ease of assessment instruments developed to be implemented in learning evaluation.
	9	Ease of assessment instruments developed for administration
Language	10	Use the Indonesian language according to the General Guidelines for Indonesian Spelling (PUEBI)
	11	Item does not contain the term "local language"
	12	Items arranged communicatively
	13	The expressions used are unambiguous

Appendix 3

Product/ Instrument