EduChemia Vol.10, No.1, 2025, p. 71-83
Jurnal Kimia dan Pendidikan e-ISSN 2502-4787

ESTIMATION OF STUDENTS' CHEMISTRY ABILITY BASED ON DICHOTOMOUS AND POLYTOMOUS DATA

Kriswantoro[1*], Lucya Wulandari[2], Suharli AJ[3]

[1]*Chemistry Education Study Program, Faculty of Teacher Training and Educational Sciences, Universitas Jambi, Indonesia*
[2]*Biology Education Study Program, Faculty of Teacher Training and Educational Sciences, Universitas Jambi, Indonesia*
[2]*Physics Education Study Program, Faculty of Teacher Training and Educational Sciences, Universitas Jambi, Indonesia*

E-mail*: kriswantoro18@unja.ac.id

Received: 26 Februari 2025. Accepted: 20 Mei 2025. Published: 31 Mei 2025

DOI: 10.30870/educhemia.v10i1.31224

Abstract: This research aims to evaluate the quality of an instrument used to measure the chemistry proficiency of senior high school students in Yogyakarta city in response to high order thinking test questions. The evaluation employs the item response theory (IRT) approach, which considers both dichotomous and polytomous data types. The assessment involved analyzing three fundamental assumptions: unidimensionality, local independence, and invariance. The unidimensionality assumption was examined through factor analysis, revealing that the first factor contributed the most to both dichotomous and polytomous data. Local independence was tested using variance-covariance matrix analysis, which revealed that the covariance values between ability intervals were minimal or nearly zero, thereby fulfilling this assumption. Furthermore, invariance testing was conducted on both item parameters and student ability parameters. The results indicated a high correlation (0.893) for item parameters and a moderate correlation (0.761) for ability parameters, confirming that the invariance assumption was satisfied. The students' ability levels were categorized into five levels, ranging from very low to very high. The analysis revealed that dichotomous data covered a broader range of abilities than polytomous data did. Overall, the study confirms that the instrument used is valid, reliable, and aligns with the core assumptions of IRT. These findings have significant implications for enhancing the quality of chemistry assessment tools using the IRT approach, ensuring that test questions are more accurate and dependable. These insights can help teachers design instructional methods that promote higher-order thinking skills (HOTSs), encourage the broader application of IRT in assessment, and support data-driven educational policies. Moreover, students gain improved training in analytical thinking and problem solving, leading to a deeper understanding of chemistry.

Keywords: Item response theory, unidimensionality, local independence, parameter invariance, chemical ability

Abstrak: Penelitian ini bertujuan untuk menguji kualitas instrumen pengukuran kemampuan kimia siswa SMA Negeri di Kota Yogyakarta dalam menjawab soal tes berpikir tingkat tinggi menggunakan pendekatan Teori Respons Butir dilihat dari jenis data dikotomus dan politimus. Pengujian dilakukan melalui analisis terhadap tiga asumsi utama, yaitu unidimensi, independensi lokal, dan invariansi parameter. Asumsi unidimensi diuji menggunakan analisis faktor, di mana hasil menunjukkan bahwa faktor pertama memberikan kontribusi terbesar pada data dikotomus dan politomus. Asumsi independensi lokal diuji melalui analisis matriks varians-kovarians, dengan hasil bahwa nilai kovarians antarinterval kemampuan sangat kecil atau mendekati nol, sehingga asumsi ini terpenuhi. Selanjutnya, pengujian invariansi parameter dilakukan untuk parameter butir dan kemampuan siswa. Hasil menunjukkan korelasi tinggi (0,893) untuk parameter butir dan korelasi moderat (0,761) untuk parameter kemampuan, mengindikasikan bahwa kedua asumsi invariansi terpenuhi. Data hasil pengukuran kemampuan siswa dikategorikan dalam bentuk predikat sangat rendah hingga sangat tinggi. Analisis menunjukkan bahwa data dikotomus memiliki rentang kemampuan yang lebih luas dibandingkan data politomus. Penelitian ini berimplikasi pada peningkatan kualitas instrumen asesmen kimia dengan pendekatan Teori Respons Butir (IRT), memastikan soal lebih valid dan reliabel. Hasilnya membantu guru merancang pembelajaran yang mendukung berpikir tingkat tinggi (HOTS), meningkatkan pemanfaatan IRT dalam evaluasi, serta mendukung kebijakan pendidikan berbasis data. Selain itu, siswa lebih terlatih dalam analisis dan pemecahan masalah, sehingga pemahaman kimia semakin mendalam.

Kata kunci: Teori Respons Butir, unidimensi, independensi lokal, invariansi parameter, kemampuan kimia.

INTRODUCTION

Chemistry education plays a crucial role in preparing skilled human resources for the challenges of industrialization and globalization. To achieve this, it must produce students with competencies in logical, analytical, and innovative thinking, problem-solving, technology mastery, and adaptability. Collaboration, particularly from teachers, is essential, making pedagogical competence a key skill for educators.

Pedagogical competence refers to a teacher's ability to oversee student learning (Kurniawan & Astuti, 2017). Teacher pedagogical competence encompasses the ability to conduct evaluations of the learning process and its outcomes. This involves (1) understanding the basics of assessment and evaluating learning outcomes; (2) identifying key elements relevant to the subject being taught; (3) establishing procedures for assessment and evaluation; (4) creating assessment tools; (5) implementing ongoing assessments using various tools; (6) interpreting assessment results for different objectives; and (7) conducting evaluations of the process and learning outcomes (Permendiknas No. 16, 2007). Based on the obligations stipulated in the Permendiknas, in the context of chemistry education, test developers, especially chemistry teachers, need to measure the abilities, learning achievements, attitudes, interests, and other characteristics of students.

Evaluation is a crucial element in education. Within the education subsystem, evaluation serves a significant purpose for educators in assessing the effectiveness of the learning process. A test is one type of evaluation that is utilized to gauge the level of student capability indirectly. A test is a method employed to discover or assess something in an environment governed by established procedures and regulations (Umami et al., 2021).

The item response theory (IRT) method provides an alternative way to analyze tests based on two main principles: relativity and probability. The principle of relativity emphasizes that the key measurement unit is the connection between student ability and item difficulty rather than just the student or item alone. In item response theory (IRT), measurement focuses not only on student ability (βn) or problem difficulty (δi) separately but also on the difference between the two (βn - δi). If βn is greater than δi, students tend to answer correctly, whereas if βn is smaller than δi, the problem is considered too difficult. If βn is close to δi, the chance of answering correctly is approximately 50%. This approach helps to construct a more accurate and fairer test for measuring student competence in chemistry education. In addition, a comparison between the student's ability and the difficulty of the item can also be made. If the student's ability surpasses the item's difficulty, the student's answer is likely to be correct. Conversely, if the student's ability is less difficult, the student's answer is likely to be incorrect (Keeves & Alagumalai, 1999; Retnawati, 2014).

Score estimation in item response theory relies on the likelihood of students providing the correct answers based on their proficiency for each question. In other words, the score estimated using IRT is the true score (Lalot et al., 2025). True scores can be obtained by applying scoring models.

Solving chemistry questions requires several stages that must be passed. Each stage requires mastery of the attributes that underlie the question. Attributes are described as procedures, processes, skills, or competencies that participants must have to solve the questions effectively (Gierl, 2007; Gierl et al., 2008; Roberts & Gierl, 2010). The attributes required to answer the questions may not always be mastered or applied correctly by participants. In dichotomous scoring, only those who fully master and apply all necessary attributes correctly receive a score, whereas others do not. Therefore, an innovative scoring model, such as a polytomous scoring system for multiple-choice chemistry questions, is needed.

One of the analytical frameworks of item response theory is frequently referred to as 1PL (one-parameter logistic). This mathematical framework was subsequently popularized by Ben Wright (Linacre, 2011). Drawing from raw data presented as dichotomous responses indicating correct and incorrect answers as measures of student capability, Rasch developed a model that links students to test items (Sumintono & Widhiarso, 2014). For dichotomous data, the Rasch model employs an algorithm that outlines the probabilistic expectations of item 'i' and respondent 'n'.

Dichotomous item responses involve two scoring options, specifically, a score of 1 for correct answers and a score of 0 for incorrect answers (Bond & Fox, 2007; DeMars, 2010). The dichotomous scoring framework is referred to as the dichotomous logistic framework, named for the count of parameters utilized in the framework (Hambleton, R. K. Swaminathan & Rogers, 1991). The one-parameter logistic model, also known as the Rasch model, is a model that includes parameters for item difficulty levels.

Polytomous scoring is a model of item response theory that features more than two potential answer categories. In this scenario, the polytomous scoring implemented is the partial credit model (PCM). The PCM does not necessitate that the steps for finishing the test items be in a specific order and that they do not need to possess the same level of difficulty (De Ayala, 1993). The threshold in PCM scoring from one category to the next does not always increase. PCMs are suitable for items that are evaluated in layered categories, although the difficulty levels at each stage are not automatically in order. Scores from higher categories indicate greater ability than scores from lower categories. In this research, responses to items could be accurate at certain steps but inaccurate at others. In these instances, it is advisable to use PCM scoring. In the RM or PCM framework, it is presumed that the item difficulty parameter is the sole characteristic of the item that affects participant performance.

In the initial stages of the advancement of polytomous item response theory, the most widely utilized model was the partial credit model (PCM), which is an evolution of the Rasch model. PCM is a polytomous scoring model created from the Rasch model intended for dichotomous data. The PCM posits that every item possesses the same level of discrimination power (Retnawati, 2011). This study aims to describe the chemistry abilities of Grade XI high school students and estimate their abilities via dichotomous and polytomous scoring models.

Low inputs, processes, and outputs of learning require an assessment (Rusmansyah et al., 2023). Thus, this study contributes to improving the quality of evaluation instruments in chemistry education using item response theory (IRT), ensuring that HOTS-based questions are more valid and reliable. The findings can assist teachers in identifying students' abilities and challenges through dichotomous and polytomous data, enabling them to develop more effective teaching strategies to foster critical thinking and problem-solving skills. Furthermore, the integration of IRT promotes technology-driven and statistical assessments; supports data-informed education policies; and enhances students' analytical, evaluative, and synthesis skills for a better understanding of chemistry.

METHOD

Owing to its nature, which aims to provide a description, this study falls under descriptive research employing a quantitative method. Quantitative descriptive research is a type of research that systematically, factually, and accurately describes facts, research objects, or the nature of a particular population. The participants in this research were eleventh-grade students from State Senior High Schools located in the city of Yogyakarta. The science program was based on the top, middle and bottom rankings of the National Examination scores throughout the city of Yogyakarta.

The data collection methods used in this research were conducted through documentation techniques. The documents used to obtain data in this study include the following: specifications of high-level thinking test questions for chemistry subjects, high-level thinking test sheets for chemistry subjects, answer keys for high-level thinking test subjects for chemistry subjects, and answer sheets for students taking high-level thinking test subjects for chemistry subjects in the academic year in Yogyakarta city. The test comprises 30 questions, and the study involved a total of 584 student participants. To demonstrate the assumption test of the item response theory, it was examined using the SPSS program by assessing unidimensionality. The analysis of chemistry test data uses two approaches to estimate student abilities, namely, dichotomous scoring via the Quest program and polytomous scoring via the Quest program.

The general form of the PCM is as follows (Retnawati, 2014):

$$P_{jk}(\theta) = \frac{exp \sum_{v=0}^{k}(\theta - b_{jv})}{\sum_{h=0}^{m} exp \sum_{v=0}^{k}(\theta - b_{jv})}, k = 0,1,2,\dots,m \quad (1)$$

$$\sum_{h=0}^{k}(\theta - b_{jh}) \equiv 0\, dan \sum_{h=0}^{k}(\theta - b_{jh}) \equiv \sum_{h=1}^{k}(\theta - b_{jh}) \quad (2)$$

with:

$P_{jk}(\theta)$: Probability of category k score on item j,

θ: participant's ability

m+1: number of categories of j grains,

b_{jv}: difficulty index of category k item j

RESULTS AND DISCUSSION

The item response theory approach first requires proof of the unidimensional assumption. Testing of this assumption is performed to ensure that the test instrument used only measures one type of trait. Proof of the unidimensional assumption is performed through factor analysis, and the results of the empirical analysis are presented in Table 1.

Table 1. KMO-MSA Test of Dichotomous and Politomous Data

KMO and Bartlett's Test		Dichotomous	Polytomous
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.786	.796
Bartlett's Test of Sphericity	Approx. Chi-Square	2483.336	2325.480
	Df	435	435
	Sig.	.000	.000

The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO-MSA) is a metric that evaluates the disparity between the correlation coefficient and the partial correlation coefficient. If the aggregate of the squares of the partial correlation coefficients for all the variable pairs is significantly smaller than the aggregate of the squares of the correlation coefficients, then the KMO value will approach 1. The KMO-MSA value is regarded as sufficient if it exceeds 0. 5 (Field, 2009). The findings of the research indicated that the dichotomous KMO-MSA value was 0. 786 and a polytomous KMO-MSA value of 0. 796 fell within the moderate range (Beavers et al., 2019).

The Bartlett test seeks to establish whether a relationship exists among variables in multivariate scenarios. If the variables X1, X2,. . . , Xp are independent (mutually independent), then the correlation matrix of the variables corresponds to the identity matrix (Field, 2009). The calculation outcomes indicate the importance of Bartlett's test of sphericity at 0. 000 for both dichotomous and polytomous datasets. Therefore, the criteria are fulfilled since the significance level is less than 0. 05. Given that the conditions of the KMO-MSA and Bartlett's test of sphericity have been satisfied, factor analysis may proceed.

Dichotomous data analysis revealed that there are 10 eigenvalues and that polytomous data have 10 eigenvalues that have a value greater than 1, where the first factor is the most dominant factor. This factor has an eigenvalue of 4.479 for dichotomous data and 4.360 for polytomous data. Nearly three times the eigenvalue of the second factor, with the third and following eigenvalues being nearly identical. In factor analysis, the first eigenvalue must possess the highest value (dominant) in comparison to the second, third, and any subsequent eigenvalues. This is because the variance explained is directly related to the magnitude of the eigenvalue (Field, 2009; Johnson & Wichern, 2002). Therefore, the initial factor in the factor analysis offers the most significant contribution in comparison to other factors, thereby satisfying the unidimensionality assumption.

Second, the conclusion of unidimensionality graphically uses a scatter plot. The number of important factors to the left of the elbow is understood as the number of dimensions created. The scree plot illustrated in Figures 1 and 2 indicates that an elbow has been established, with one point located to the left of the elbow (Hambleton, R K. Swaminathan & Rogers, 1985). Referring to the views of the above experts, it can be concluded that the unidimensional assumption holds true. This is more clearly illustrated in Figures 1 and 2.

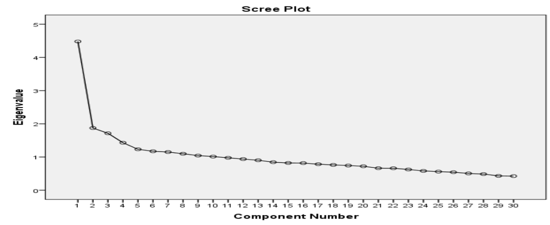

Figure 1. Dichotomous eigenvalue scree plot

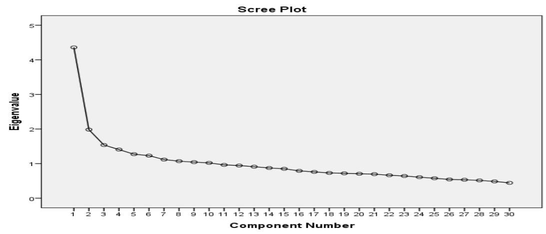

Figure 2. Polytomus eigenvalue scree plot

The local independence requirement test aims to ensure that students' abilities are independent of the test items. In other words, a student's answer to one test item does not affect the explanation of other test items. This test can be performed by analyzing the variance-covariance matrix of students' abilities that have been collected into several subgroups. The premise of local independence is satisfied if the covariance value among ability intervals is low or nearly zero (Hambleton, R. K. Swaminathan & Rogers, 1991). The subsequent section displays the outcomes of the local independence analysis, as shown in Table 2.

Table 2. Variance-covariance matrix

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_10
K_1	0.6719									
K_2	0.1044	0.0266								
K_3	0.0771	0.0175	0.017							
K_4	0.0701	0.0158	0.0154	0.0139						
K_5	0.0817	0.0162	0.0146	0.0132	0.0139					
K_6	0.0481	0.0102	0.0075	0.0068	0.0071	0.0061				
K_7	0.0507	0.0130	0.0107	0.0096	0.0094	0.0058	0.0093			
K_8	0.0578	0.0132	0.0123	0.0111	0.0109	0.0059	0.0085	0.0099		
K_9	0.0673	0.0167	0.0117	0.0106	0.0109	0.0071	0.0093	0.0090	0.0122	
K_10	0.1389	0.0376	0.0296	0.0267	0.0262	0.0162	0.0227	0.0229	0.0260	0.0787

Table 2 presents the variance-covariance matrix values between student ability groups. Before that, student abilities were grouped by sorting them from highest to lowest and then divided into 10 groups. The analysis process used to calculate the variance-covariance matrix was carried out using Excel. Overall, the components outside the main diagonal in this matrix appear minimal or nearly nil; thus, it can be inferred that the premise of local independence has been satisfied.

The parameter invariance test seeks to verify that the attributes of the test items remain the same regardless of the different groups of students who answer them. Similarly, the assessment of students' abilities within the same group will not fluctuate even if the questions they respond to differ. Therefore, there are two categories of parameter invariance: item parameter invariance and parameter invariance.

To test item parameter invariance, 584 students were divided into two groups based on their serial numbers: group I consisted of odd-numbered students, and group II consisted of even-numbered students. The item parameter estimation, in this case the level of difficulty of the test items analyzed using the Rasch model, was calculated for each group. The results of the estimation were subsequently visualized using a scatter plot and correlated. If the correlation is both positive and significant, then the assumption regarding item parameter invariance is deemed met (Retnawati, 2014).

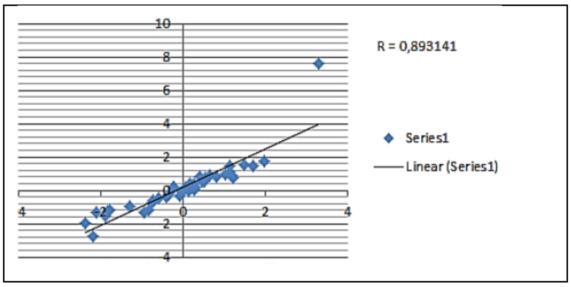

Figure 3. Item parameter invariance plot

Figure 3 shows the estimation plot used to evaluate the invariance of the item parameters. The figure indicates that the estimated values are quite aligned with a straight line, exhibiting a notably high correlation of 0. 893. Thus, it can be inferred that the assumption of invariance of item parameters has been satisfied.

To test the invariance of the ability parameter (θ), thirty questions were categorized into two subtest groups according to the item number. Subtest I consists of odd-numbered questions, whereas subtest II consists of even-numbered questions. The estimated ability parameters of each subtest were then visualized in the form of a scatter plot and correlated. If the results of the correlation are positive and high, it can be inferred that the assumption of invariance of the ability parameters has been satisfied (Retnawati, 2014).

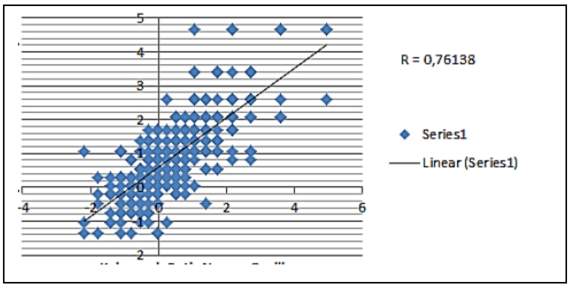

Figure 4. Scatter plot of the capability parameters

Figure 4 above displays a scatter plot of the ability parameters derived from the set of questions addressed by the students. According to the figure, the estimated values are positioned fairly near a straight line with a strong (considerable) correlation value of 0.761. The correlation value of 0.60 is a reasonable value (moderate) and is most commonly found in the field of education (Best & J.V., 1998). Thus, it can be inferred that the premise of the constancy of ability parameters has also been satisfied.

Measurement data of ability are used as a source of information to describe the chemistry ability of senior high school students in Yogyakartacity. These data are obtained through the interpretation of scores that produce values. The values can be presented either in the form of numbers or predicates. The outcomes of the analysis of scores or evaluations are utilized for multiple objectives, including enhancing the quality of education, documenting learning results, and performing similar tasks. Assessments were conducted in six schools that were classified into high, medium, and low categories according to the National Examination scores.

Students' chemistry ability can also be presented in the form of predicates ranging from very low to very high. To produce these predicates, the scores are divided into several categories. In this study, for dichotomous data, the maximum ability (θ) reached 3.13, whereas the minimum was -2.61, with an ideal average value of 0.26 and an ideal standard deviation of 0.08. Moreover, for polytomous data, the maximum ability (θ) was 1.93, and the minimum was -0.47, with an ideal average value of 0.73 and an ideal standard deviation of 0.24. The range of values is divided into five standard deviation units. The distribution of abilities from dichotomous and polytomous data is shown in Figure 5.

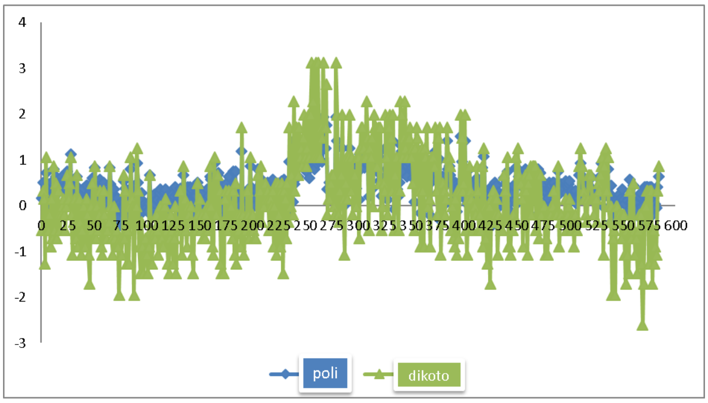

Figure 5. Distribution of dichotomous and polytomous data

As shown in Figure 5, the distribution of dichotomous and polytomous abilities indicates that for dichotomous data, the ability is greater than that for polytomous data. The highest and lowest abilities are owned by dichotomous data, whereas for polytomous data, the abilities seen do not exceed and are not lower than the data owned by dichotomous data.

The estimation of students' chemistry ability based on dichotomous and polytomous data needs to be performed because each type of data has certain characteristics and advantages that provide additional information in the analysis of student ability. Dichotomous data are used to measure student responses that are true or false, such as multiple-choice questions. These data are suitable for describing students' basic abilities. Polytomous data are used to measure responses in more than two categories, such as assessment scales or essay questions. These data provide more detailed information about students' ability levels at various levels.

By integrating dichotomous and polytomous data, the analysis becomes more comprehensive. Dichotomous data rpovide a general picture of ability, whereas polytomous data provide details about the quality of students' understanding at various levels of difficulty. Estimates from both types of data allow teachers or researchers to understand students' abilities as a whole, both in terms of basic knowledge (dichotomous) and critical or analytical thinking skills (polytomous). By comparing the results of dichotomous and polytomous data, the measurement instrument can be tested for its reliability in providing consistent ability estimates across various question forms. The result of ability estimates from both types of data provide deeper insight into students' needs and weaknesses. This can be used to design learning strategies that are more effective and tailored to the needs of each student.

Therefore, learning chemistry requires learning strategies and teachers who can direct students to understand and connect the three levels of chemical representation (Langitasari & Robandi, 2023). This study has several key implications for chemistry education, particularly in enhancing the quality of evaluation instruments. Through the application of item response theory (IRT) in analyzing high-order thinking (HOTS)-based questions, educators can develop test instruments that are more valid and reliable, ensuring a more accurate assessment of students' abilities. The findings assist teachers in identifying students' difficulties and strengths in answering questions based on dichotomous and polytomous data, enabling the development of more effective learning strategies to enhance critical thinking and problem-solving skills in chemistry. The use of IRT-based analysis allows schools and educators to implement more advanced evaluation methods beyond conventional approaches, promoting the integration of technology and statistical techniques in learning assessments. Additionally, the study provides a foundation for education policy development, particularly in designing curricula that emphasize HOTS-based evaluation. By refining assessment instruments, students can improve their analytical, evaluative, and synthesis skills, which are essential for a deeper understanding of chemical concepts.

CONCLUSION

The study confirms that the test instrument meets the basic assumptions of item response theory (IRT), including unidimensionality, local independence, and invariance. The results show that the instrument provides valid and reliable measures of student ability, with dichotomous data offering a wider ability range than polytomous data. These findings have significant implications for chemistry education, helping teachers design more effective learning strategies, promoting critical thinking, and supporting data-driven educational policies. The application of IRT also encourages the use of technology and statistical methods to enhance students' analytical and problem-solving skills, leading to a deeper understanding of chemistry.

REFERENCES

Beavers, A., Lounsbury, J., Richards, J., Huck, S., Skolits, G., & Esquivel, S. (2019). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, *18*(3), 1–13.

Best, J. W., & J.V., K. (1998). *Educational research (8th ed.)*. Allyn & Bacon.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences. (2nd Ed.)*. Lawrence Erlbaum Associates Publishers.

De Ayala, R. J. (1993). Methods, Plainly Speaking: An Int roduct ion to Polytomous Item Response Theory Models. *Measurement and Evaluation in Counseling and Development*, *25*, 172–189.

DeMars, C. (2010). *Item Response Theory*. Oxford University Press, Inc.

Field, A. (2009). *Discovering statistics using SPSS (3rd ed.).* Sage Publication Ltd.

Gierl, M. J. (2007). Making Diagnostic Inferences about Cognitive Attributes Using the Rule-space Model and Attribute Hierarchy Method. *Journal of Educational Measurement*, *45*(1), 325–340.

Gierl, M. J., Zheng, Y., & Cui., Y. (2008). Using the Attribute Hierarchy Method to Identify and Interpret Cognitive Skills that Produce Group Differences. *Journal of Educational Measurement*, *45*(1), 65–89.

Hambleton, R. K. Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory.* Sage Publications.

Hambleton, R K. Swaminathan, H., & Rogers, H. J. (1985). *Item Response Theory: Principles and Applications*. Kluwer-Nijjhoff Publish.

Heale, R., & Twycross, A. (2018). What is a case study? *Evidence-Based Nursing*, *21*(1), 7–8. doi: 0.1136/eb-2017-102845

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis.* Prentice-Hall, Inc.

Keeves, J. ., & Alagumalai, S. (1999). *New Appoaches to Measurement*. Pergamon.

Kurniawan, A., & Astuti, A. P. (2017). Deskripsi Kompetensi Pedagogik guru dan Calon Guru Kimia SMA Muhammadiyah 1 Semarang. *Seminar Nasional Pendidikan, Sains Dan Teknologi*, 1–7.

Lalot, F., Räikkönen, J., & Ahvenharju, S. (2025). An item response theory approach to measurement in environmental psychology: A practical example with environmental risk perception. *Journal of Environmental Psychology*, *101*(January). doi: 10.1016/j.jenvp.2025.102520

Langitasari, I., & Robandi, B. (2023). Analysis of The Student's Ability to Interconnect Macro-Submicro-Symbolic Representation on Electrolyte Solution Concept. *EduChemia (Jurnal*

Kimia Dan Pendidikan), *8*(1), 84. doi: 10.30870/educhemia.v8i1.19974

Linacre, J. M. (2011). *A User's guide to WINSTEPS Ministeps; Rasch-model Computer Program. Program Manual 3.73*.

Retnawati, H. (2011). Mengestimasi Kemampuan Peserta Tes Uraian Matematika Dengan Penskoran Politomus dengan Generalized Partial Credit Model. *Seminar Nasional Penelitian, Pendidikan Dan Penerapan MIPA 2011*, 53–62.

Retnawati, H. (2014). *Teori Respon Butir dan Penerapannya*. Parama Publishing.

Roberts, M. R., & Gierl, M. J. (2010). Developing Score Reports for Cognitive Diagnostic Assessment. *Educational Measurement*, *29*(3), 25–38.

Rusmansyah, R., Effendi, A. M., Mahdian, M., & Aufa, M. N. (2023). Scientific Critical Thinking Model to Improve Scientific Literacy and Self-Regulation Skills in Buffer Solution Learning. *EduChemia (Jurnal Kimia Dan Pendidikan)*, *8*(2), 130. doi: 10.30870/educhemia.v8i2.15988

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. trim Komunikata Publishing House.

Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan instrumen tes untuk mengukur higher order thinking skills (HOTS) berorientasi programme for international student asessment (PISA) pada peserta didik. *JP3M (Jurnal Penelitian Pendidikan Dan Pengajaran Matematika)*, *7*(1), 57–68. doi: 10.37058/jp3m.v7i1.2069