Jurnal Pijar MIPA
https://jurnalfkip.unram.ac.id/index.php/JPM

ISSN 1907-1744 (Print)
ISSN 2460-1500 (Online)

Design of a Chemical Literacy Test Instrument on Atomic Structure Material for Senior High School Students

Siti Faizah, Yerimadesi*, Nofri Yuhelman, Iryani

Department of Chemistry Education, Faculty of Mathematics and Natural Science, Padang State University, Padang, Indonesia
*e-mail: yeri@fmipa.unp.ac.id

Received: January 15, 2026. Accepted: February 17, 2026. Published: February 26, 2026

Abstract: The low level of chemical literacy in Indonesia, reflected in PISA 2022 scores below the average, is driven by the limited use of literacy-based assessments and restricted access to proper evaluation instruments. This study aims to develop and validate a chemical literacy test instrument specifically designed for atomic structure material for Senior high School students, utilizing the Rasch model for rigorous psychometric analysis. Unlike previous studies that often integrate multiple chemistry topics, this research provides a specialized focus exclusively on atomic structure, featuring five contextual narratives and one infographic across seven essay items to capture deep conceptual understanding. This Research and Development (R&D) employed purposive sampling of students at SMA N 7 Padang. Data were analyzed using Content Validity Ratio (CVR) for logical validity and Rasch modelling via Ministep software for empirical validation. The Rasch model was selected over classical test theory to ensure linear measurement of both item difficulty and student ability. SME evaluations confirmed the logical validity of all items. Empirical analysis showed that although some items exhibited higher outfit values, they were retained because of strong point-measure correlations with the measured construct. Item reliability was categorized as good, person reliability as fair, and Cronbach's alpha was categorized as good. The instrument effectively categorized item difficulty and student discrimination power into three ability levels (high, medium, and low). The resulting seven-item instrument is valid, reliable, and practically feasible for educators to integrate into formative assessments to measure students' chemical literacy in atomic structure.

Keywords: Atomic Structure; Chemical Literacy; Instrument Test; Rasch Model.

Introduction

Chemical literacy is an essential component of scientific literacy, vital in today's educational landscape due to the close relationship between chemistry and daily life. It is imperative that chemical literacy be taught to students in the 21st century, where science and technology are advancing rapidly [1]. Chemical literacy encompasses facts and phenomena occurring globally, such as global warming, energy crises, and environmental degradation [2]. Despite its paramount importance, not all students effectively comprehend chemical literacy. One of the current challenges in Indonesian education is the low level of general literacy, which includes chemical literacy.

This low level of chemical literacy is evident in PISA (Programme for International Student Assessment) scores. PISA evaluates scientific literacy outcomes (including chemical literacy) among 15-year-old students every three years, organized by the Organization for Economic Cooperation and Development (OECD) [3]. Indonesia has participated in this program since 2000. From 2000 to 2022, Indonesia has consistently recorded PISA scores below the international average. In the 2022 PISA results, Indonesia ranked 68th in scientific literacy out of 81 participating countries, earning a score of 383, while the OECD average score reached 485 [4].

Measuring a student's comprehensive level of chemical literacy is no simple task. One cause for the low chemical literacy outcomes is the insufficient use of chemical literacy-based questions in the learning process [5]. Assessment instruments commonly used in schools often emphasize low-level cognitive aspects, such as rote memorization of concepts, definitions, or routine calculations. Consequently, these instruments are less capable of exploring and measuring students' abilities to apply chemical knowledge to daily life, evaluate information, or integrate various concepts to solve complex problems[6]. The difficulty in finding chemical literacy instruments means that many teachers have not yet utilized literacy-type questions to measure their students' abilities, and are looking for guides or instrument examples [7].

The lack of use and the difficulty in accessing chemical literacy questions directly affect students' literacy skills. Strong chemical literacy is a foundational requirement for students to understand fundamental concepts, including atomic structure. Chemical literacy is not limited to memorizing formulas and definitions; it includes the ability to interpret information, think critically, and apply chemical concepts across various contexts[8]. In the study of atomic structure, chemical literacy enables students to analyze atomic models over time, understand how the discovery of subatomic particles changed the perspective on atoms, and link electron configurations to elemental properties [9], [10], [11]. Without chemical literacy, students may struggle to visualize atoms, understand the abstract concept of quantum numbers, or relate atomic structure to the formation of

How to Cite:
S. Faizah, Y. Yerimadesi, N. Yuhelman, and I. Iryani, "Design of a Chemical Literacy Test Instrument on Atomic Structure Material for Senior High School Students", *J. Pijar.MIPA*, vol. 21, no. 1, pp. 164–170, Feb. 2026. https://doi.org/10.29303/jpm.v21i1.11447

chemical bonds [12]. Therefore, developing students' chemical literacy is crucial to fostering a deep understanding of atomic structure, which, in turn, supports their future success in learning more complex chemical concepts.

The topic of Atomic Structure is one of the chemistry subjects included in the Learning Outcome of SK BSKAP No. 32 of 2024. This topic is taught in Phase E of Senior High School. Atomic structure is the fundamental framework for understanding advanced chemical concepts such as stoichiometry, chemical reactions, equations, and bonding. The curriculum covers the formation of atoms from the beginning of the universe and their evolution into the physical structure of the atom, through the discovery of subatomic particles. Furthermore, atomic structure explains how macroscopic properties (those visible to the eye) are dependent on microscopic properties (molecular or atomic levels; not physically visible) [13].

Regarding students' comprehension of atomic structure, the results were significantly low [14]. This outcome is influenced by several factors, notably that chemistry instruction on this subject in senior high schools is frequently dominated by conventional teacher-centered methods. Such approaches often fail to accommodate diverse student learning styles, leading to low learning motivation and a lack of active engagement. Consequently, this negatively impacts both academic performance and levels of chemical literacy. To address these challenges, there is a pressing need for innovation in pedagogical strategies and the development of relevant assessment instruments. Furthermore, many students encounter difficulties when solving problems related to atomic structure, particularly in distinguishing between various atomic theories [15]. As the subject matter primarily comprises theoretical and abstract concepts, it remains inherently difficult for students to grasp.

Developed a chemical literacy instrument for atomic structure and nanotechnology in Phase E; the scope of that instrument remains broad. By covering both atomic structure and nanotechnology, the depth of exploration for each specific topic is somewhat limited. The novelty of the present study lies in its exclusive and intensive focus on atomic structure, encompassing the history of atomic models, subatomic particles, atomic classification, and electron configurations. This specialized approach is expected to provide a more comprehensive measurement of each sub-topic within the atomic structure material. Furthermore, by utilizing the Rasch model analysis, this research addresses the methodological gaps of previous studies, ensuring a more linear and objective evaluation of student literacy levels compared to broader, multi-topic assessments [16].

The designing of valid and reliable test instruments is vital for accurately measuring students' chemical literacy levels [17]. These instruments are expected to provide a clearer picture of student performance. Furthermore, they serve not only as summative assessment tools but also as formative assessments to monitor student learning progress. Therefore, this study aims to bridge the existing literature gap by developing a chemical literacy test instrument with a deep, exclusive focus on atomic structure and validating it through Rasch modelling. By employing a specialized content approach and rigorous psychometric validation, this instrument is expected to provide educators with a more accurate and objective measurement tool to comprehensively map students' chemical literacy profiles.

Research Methods

This study is a Research and Development (R&D) project, which examines the design and development of learning products, such as test instruments, and evaluates their effectiveness [18]. The design of this test instrument refers to the ten stages of instrument development proposed by Xiufeng Liu et al. (2020), utilizing the Rasch Model [19]. This analytical approach provides more precise measurements and accurate information regarding both the test instrument's quality and the students' abilities. The development and validation processes were conducted at the Chemistry Education Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, while the field testing was carried out at SMA N 7 Padang. The research instruments utilized in this study consist of validation sheets and the newly developed chemical literacy test instrument. The data collected are quantitative, derived from the validity assessment scores provided by expert validators and the results of the field trials conducted with students.

The data analysis techniques employed in this study include validity and reliability testing. Logical validity was assessed through validation instruments evaluated by Subject Matter Experts (SMEs). The resulting validity data were analyzed using the Content Validity Ratio (CVR) developed by Lawshe (1975) to quantify expert agreement regarding the essentiality of specific items [20]. According to Lawshe's criteria, if more than half of the SMEs deem an item relevant, that item is considered to possess sufficient content validity.

$$\mathrm{CVR} = \frac{n_e - (\mathrm{N}/2)}{\mathrm{N}/2}$$

Information:

n_e = the number of validators indicating an item is "essential" or relevant

N = the total number of validators

The assessment criteria for the acceptance of test instruments based on the CVR scale range from +1 to -1. A higher CVR value indicates that an item is more essential, reflecting greater validity. For a panel of five validators, the minimum threshold for a CVR value to be accepted and considered valid is 0.736 [20].

Once the instrument was declared valid through the expert validation process, the chemical literacy instrument was field-tested with students at SMA N 7 Padang during the 2025/2026 academic year (specifically the July–December 2025 semester). The resulting data were analyzed quantitatively, encompassing empirical validity, reliability, item difficulty, and discrimination power. Empirical validity, reliability, item difficulty, and discrimination power were analyzed using Rasch modelling facilitated by the Ministep software. The criteria for empirical validity were determined by assessing item fit using Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point Measure Correlation (Pt Mean Corr). According to Rasch modelling standards, an item is categorized as 'fit' if it meets the following requirements: an Outfit MNSQ value between 0.5

and 1.5, an Outfit ZSTD value between -2.0 and +2.0, and a Pt Mean Corr value between 0.4 and 0.85 [21].

The results of the reliability analysis can be observed in the summary statistics section of the Ministep application, specifically focusing on person reliability, item reliability, and Cronbach's alpha coefficients. The criteria for determining the quality of person and item reliability are categorized according to the standards presented in Table 1 below [21].

Table 1. Criteria for reliability interpretation

Value	Criteria
< 0.67	Weak
0.67 – 0.80	Fair
0.80 – 0.90	Good
0.90 – 0.94	Very good
> 0.94	Excellent

Furthermore, the overall reliability, which represents the interaction between items and individuals, is measured by Cronbach's alpha. The criteria for this value are categorized as shown in Table 2 [21].

Table 2. Criteria of Alpha Cronbach

Value	Criteria
< 0.5	Poor
0.5 – 0.6	Bad
0.6 – 0.7	Fair
0.7 – 0.8	Good
> 0.8	Very good

The difficulty level of the test items is determined using the difficulty index. A high-quality instrument consists of items that are neither excessively easy nor overly difficult [22]. Within the Ministep output menu, item difficulty is determined by selecting Item Measure. The difficulty levels are classified into four categories based on logit values: items with a logit value exceeding +1SD are categorized as Very Difficult, values between 0.0 and +1SD are categorized as Difficult, values between 0.0 and -1SD are categorized as Easy, and values below -1SD are categorized as Very Easy [21].

Following the assessment of item difficulty, item discrimination can be evaluated through the separation value. Discrimination power refers to an item's ability to distinguish between groups based on the measured aspects, in accordance with the actual differences within those groups. This value can be found in the summary statistics within the Ministep Output Table menu. To evaluate discrimination, the individual ability levels are examined. Furthermore, a specific equation used to categorize these groups more precisely is referred to as strata separation [21].

$$\mathrm{H} = \frac{[(4 \times \text{separation}) + 1]}{3}$$

Results and Discussion

Based on the research conducted, the product developed is a chemical literacy instrument comprising 7 essay items on atomic structure for Senior High School students. Before the development process begins, it is essential to define the instrument's specific objectives. This step is crucial as it serves as a guide throughout the development phase. Specifically, this research seeks to develop a chemical literacy instrument capable of measuring students' chemical literacy skills within the topic of atomic structure. Furthermore, this instrument is intended to serve as an evaluative tool for teachers to assess students during the learning process. The target population for this study consists of senior high school students.

In the subsequent stage, the Learning Outcomes were analysed to derive the Learning Objectives. Based on the results of the learning objectives analysis, a test blueprint was constructed, encompassing the item indicators and the designated test formats. The test instrument was then designed in accordance with this blueprint. The resulting design consists of six passages, each with seven sub-questions. The finalized design of the chemical literacy items is presented in the table below.

Table 3. Design of chemical literacy test items

Passage 5: Cyanide

Please read the following passage carefully!

Cyanide (Figure 6) is a chemical compound consisting of carbon and nitrogen atoms.**(Context)** Carbon atoms contain 6 protons and have three naturally occuring atomic masses: C^{12}, C^{13}, and C^{14}. Nitrogen atoms, which contain 7 protons, also have two naturally occuring atomic masses: N^{14} and N^{15}. **(Content Knowledge)**

Figure 6. Cyanide

In their pure elemental form, Carbon (C) and Nitrogen (N) are not hazardous. However, when combined into a compound such as cyanide, they become a notorious and highly lethal poison. **(HOLS)** Cyanide is extremely toxic and potentially fatal due to its ability to interfere with the respiratory process. This damage occurs rapidly in organs that require high oxygen levels, such as the brain and heart, leading to severe symptoms like seizures, loss of conciousness, lungdamage, and in large doses, resulting in death within minutes. **(Affective)**

Please answer the following question carefully!

Based on the passage above, identify which atoms are categorized as isotopes, isobars, and isotones from the elements composing the cyanide compound. Explain your answer by correlating them with the number of subatomic particles, atomic numbers, and mass numbers!

Table 3 presents the design of the chemical literacy test instrument, which incorporates three to four aspects of chemical literacy (content, context, HOLS, and affective). The content aspect explains the chemical science in the passage, where chemistry provides knowledge to understand materials that encompass core chemical ideas [23]. In Table 3, the content aspect concerns the atoms of the elements that compose the cyanide compound, including carbon (C) and nitrogen (N). The context aspect involves applying chemical concepts to real-life situations. The context presented in Table 3 is cyanide, a naturally occurring chemical compound. The HOLS (Higher Order Learning Skills) aspect emphasizes students' thinking abilities to explain and identify questions based on the information provided; this is linked to the content and context aspects previously presented. The HOLS aspect in Table 3 provides information on the constituent elements of the cyanide compound, and students are expected to identify them using the given information. The affective aspect involves responding to the facts about cyanide presented in the information, enabling students to maintain a rational perspective on its use.

Table 4. Chemical literacy grading criteria

<table>
<tr><th>Answer Description</th><th>Chemistry Literacy Level</th><th>Score</th></tr>
<tr><td>No answer/Incorrect answer</td><td>Scientific illiteracy</td><td>0</td></tr>
<tr><td>Isotopes: Atoms possessing the same atomic number but differing in mass numbers.
Isobars: Atoms from different elements sharing the same mass number.
Isotones: Atoms from different elements sharing an identical number of neutrons</td><td>Nominal scientific literacy</td><td>1</td></tr>
<tr><td>Isotopes: Atoms possessing the same atomic number but differing in mass numbers.
Example: the atoms of C^{12}, C^{13}, C^{14} are isotopes
The atoms of N^{14} and N^{15} are isotopes

Isobars: Atoms from different elements sharing the same mass number.
Example: the atoms of C^{14} and N^{14} are isobars

Isotones: Atoms from different elements sharing an identical number of neutrons
Example: the atoms of C^{13} and N^{14} are isotones
The atoms of C^{14} dan N^{15} are isotones</td><td>Functional scientific literacy</td><td>2</td></tr>
<tr><td>Isotopes: Atoms possessing the same atomic number but differing in mass numbers.
Example: the atoms of C^{12}, C^{13}, C^{14} are isotopes beacuse they have the same number of protons, which is 6.
<table>
<tr><th>Atom</th><th>Atomic number</th><th>Mass number</th><th>Proton</th></tr>
<tr><td>C^{12}</td><td>6</td><td>12</td><td>6</td></tr>
<tr><td>C^{13}</td><td>6</td><td>13</td><td>6</td></tr>
<tr><td>C^{14}</td><td>6</td><td>14</td><td>6</td></tr>
</table>
The atoms of N^{14} and N^{15} are isotopes because they have the same number of protons, 7.
<table>
<tr><th>Atom</th><th>Atomic number</th><th>Mass number</th><th>Proton</th></tr>
<tr><td>N^{14}</td><td>7</td><td>14</td><td>7</td></tr>
<tr><td>N^{15}</td><td>7</td><td>15</td><td>7</td></tr>
</table>
Isobars: Atoms from different elements sharing the same mass number.
Example: the atoms of C^{14} and N^{14} are isobars beacuse they have the same mass numbers, which is 14.
<table>
<tr><th>Atom</th><th>Atomic number</th><th>Mass number</th><th>Proton</th></tr>
<tr><td>C^{14}</td><td>6</td><td>14</td><td>6</td></tr>
<tr><td>N^{14}</td><td>7</td><td>14</td><td>7</td></tr>
</table>
Isotones: Atoms from different elements sharing an identical number of neutrons
Example: the atoms of C^{13} and N^{14} are isotones because they have the same number of neutrons, 7.
<table>
<tr><th>Atom</th><th>Atomic number</th><th>Mass number</th><th>Neutron</th></tr>
<tr><td>C^{13}</td><td>6</td><td>13</td><td>7</td></tr>
<tr><td>N^{14}</td><td>7</td><td>14</td><td>7</td></tr>
</table>
The atoms of C^{14} and N15 are isotones because they have the same number of neutrons, 8.
<table>
<tr><th>Atom</th><th>Atomic number</th><th>Mass number</th><th>Neutron</th></tr>
<tr><td>C^{14}</td><td>6</td><td>14</td><td>8</td></tr>
<tr><td>N^{15}</td><td>7</td><td>15</td><td>8</td></tr>
</table>
</td><td>Conceptual scientific literacy</td><td>3</td></tr>
</table>

Table 4 presents the assessment rubric derived from the passage in Table 3. There are four levels of chemical literacy that students can achieve when answering this question. The scientific illiteracy level indicates that students are unable to correlate with or respond to the questions; they fail to understand the material concepts and context, and they lack the cognitive ability to identify the questions provided. This level is assigned a score of 0 because no correct answers are obtained. The nominal scientific literacy level indicates that students recognize the concepts and context of the chemical material, but cannot provide a meaningful explanation, leading to misconceptions where students only memorize concepts or terminology. At this level, students only recognize the terms isotopes, isobars, and isotones but may experience misconceptions or errors in their classification, thus receiving a score of 1. The functional scientific literacy level indicates that students have the ability to define concepts correctly based on their understanding, although their comprehension remains limited, which is equivalent to the comprehension level (C2) of Bloom's Taxonomy [24]. At this level, students can explain the terms and categorize atoms into their respective groups, earning a score of 2. The final stage in this rubric is the conceptual scientific literacy level, which indicates that students understand scientific concepts conceptually and can integrate as well as organize the information provided in the passage rather than relying on rote memorization. At this level, students understand the classification of isotopes, isobars, and isotones within the atoms composing the cyanide compound. Beyond knowing the theory, students at this level also understand the underlying rules of these classifications, for instance, that isotopes are atoms with the same number of protons and are therefore awarded a score of 3.

Validation of the test instrument was conducted to determine its accuracy in measuring the intended constructs [25]. Validity data were obtained through an analysis of the instrument validation conducted by Subject Matter Experts (SMEs), consisting of five validators, all chemistry lecturers from the Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang (FMIPA UNP). The chemical literacy validation instrument encompasses four dimensions: content, construction, language, and graphics. The results of the item analysis are detailed in Table 5.

Based on the validity analysis of the chemical literacy test items presented above, all seven items are declared valid, as the obtained CVR values exceed the critical threshold (CVR > 0.736). Specifically, items 5 and 6 achieved a CVR of 0.93, item 7 reached 0.86, and items 1, 2, 3, and 4 achieved a CVR of 0.81. Items are considered valid when their validity score exceeds the minimum CVR threshold, indicating they meet the requirements for content validity [26]. These results signify that the instrument is ready for the subsequent research stage, following minor revisions based on the validators' suggestions and feedback.

Table 5. CVR analysis results for chemical literacy items

Test item	CVR value	Interpretation
1	0.81	Valid
2	0.81	Valid
3	0.81	Valid
4	0.81	Valid
5	0.93	Valid
6	0.93	Valid
7	0.86	Valid

The designed test instrument was administered to a sample of students to collect data for Rasch analysis. The instrument was tested on 35 tenth-grade students at SMA N 7 Padang who had completed the atomic structure topic. The resulting student response scores were analyzed using the Rasch model facilitated by the Ministep application. The evaluated item qualities included validity, reliability, difficulty level, and discrimination power.

Item's validity

Following the field testing, the data were collected and analyzed. The next phase involved empirical validity testing. Empirical validity refers to item validity derived from actual student performance data [27]. In this study, item validity was assessed through the Item Fit Order, utilizing three fit criteria: Outfit Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point Measure Correlation (Pt Measure Corr). An item is considered valid if it meets at least two of these three fit criteria; it is not mandatory to satisfy all three [28]. The results of the item analysis are presented in the table below.

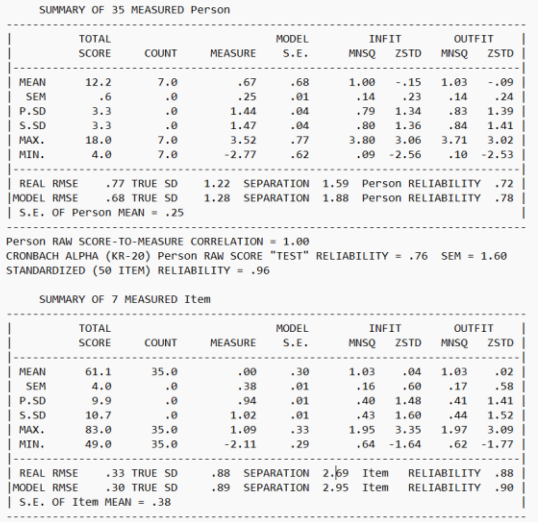
Table 6. Analysis results of empirical validity

Number of Items	Outfit		Pt Measure Corr	Interpretation
	MNSQ	ZSTD		
5	1.97	3.09	.61	Accepted
2	1.13	.62	.66	Accepted
7	.91	-.33	.59	Accepted
6	.89	-.42	.77	Accepted
1	.88	-.45	.58	Accepted
3	.84	-.62	.71	Accepted
4	.62	-.1.77	.57	Accepted

As shown in Table 6, six of the seven items, specifically 2, 7, 6, 1, 3, and 4, met all three item-fit criteria, indicating they are feasible and effective for measurement. Meanwhile, one item (Item 5) satisfied only one of the fit criteria, while the other two fell outside the acceptable range. However, Item 5 remains acceptable because its Pt Mean Corr value still falls within the favorable category. Item 5 only satisfied the Pt Mean Corr criterion, whereas the Outfit MNSQ and ZSTD values exceeded the required thresholds. The irregular response patterns from both high-ability and low-ability students (noise) caused the outfit results for Item 5 to exceed the criteria. In Rasch modelling, a 'fit' item must exhibit a consistent response pattern, where students with high ability (high total scores) tend to score higher on that item, and students with low ability score lower [28]. Nevertheless, this item was retained in the test instrument. This decision was based on a PT Mean Corr value of 0.61, indicating that Item 5 has very strong discriminatory power and aligns with the measured construct. Items with high outfit values but strong Point Measure Correlation (0.61) are categorized as 'unproductive for measurement' but not 'misleading.' This implies that although the item contains some unexpected data (noise), it still provides a functional contribution to ranking respondents based on their abilities [29]. Consequently, based on the empirical validity analysis, all items are deemed valid and suitable for measurement.

Reliability

In the context of Rasch measurement, reliability is assessed through Person Reliability and Item Reliability indices, which indicate the replicability of person and item orderings. Reliability refers to the extent to which an instrument provides consistent and stable results when the same measurement is conducted repeatedly (or under similar conditions) [30]. The reliability results of the chemical literacy test instrument on atomic structure, obtained from the analysis using Ministep, are presented in Figure 1.



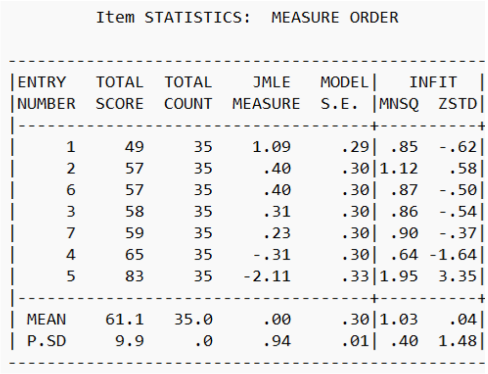
```
     SUMMARY OF 35 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      12.2        7.0        .67     .68      1.00   -.15   1.03   -.09 |
|  SEM        .6         .0        .25     .01       .14    .23    .14    .24 |
| P.SD       3.3         .0       1.44     .04       .79   1.34    .83   1.39 |
| S.SD       3.3         .0       1.47     .04       .80   1.36    .84   1.41 |
| MAX.      18.0        7.0       3.52     .77      3.80   3.06   3.71   3.02 |
| MIN.       4.0        7.0      -2.77     .62       .09  -2.56    .10  -2.53 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .77 TRUE SD    1.22  SEPARATION  1.59  Person RELIABILITY  .72 |
|MODEL RMSE    .68 TRUE SD    1.28  SEPARATION  1.88  Person RELIABILITY  .78 |
| S.E. OF Person MEAN = .25                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .76  SEM = 1.60
STANDARDIZED (50 ITEM) RELIABILITY = .96

     SUMMARY OF 7 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      61.1       35.0        .00     .30      1.03    .04   1.03    .02 |
|  SEM       4.0         .0        .38     .01       .16    .60    .17    .58 |
| P.SD       9.9         .0        .94     .01       .40   1.48    .41   1.41 |
| S.SD      10.7         .0       1.02     .01       .43   1.60    .44   1.52 |
| MAX.      83.0       35.0       1.09     .33      1.95   3.35   1.97   3.09 |
| MIN.      49.0       35.0      -2.11     .29       .64  -1.64    .62  -1.77 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .33 TRUE SD     .88  SEPARATION  2.69  Item   RELIABILITY  .88 |
|MODEL RMSE    .30 TRUE SD     .89  SEPARATION  2.95  Item   RELIABILITY  .90 |
| S.E. OF Item MEAN = .38                                                     |
-------------------------------------------------------------------------------
```

Figure 1. Summary statistic

Figure 1 presents the summary statistics for the reliability analysis and Cronbach's alpha. Cronbach's Alpha is utilized to measure the overall reliability by examining the interaction between items and individuals. A Cronbach's Alpha value between 0.7 and 0.8 indicates that the interaction between items and respondents is reliable and falls into the 'good' category [30]. In this study, the Cronbach's Alpha value was 0.76, which is categorized as good. The item reliability value of 0.88 indicates good item reliability. Furthermore, the person's reliability value was 0.72, which falls into the 'fair' category. Thus, it can be concluded that the test items are appropriate for the students and that their response patterns are reasonably consistent, which aligns with the findings of [16].

Difficulty level

Item difficulty level indicates whether a test item falls into the easy, moderate, or difficult category. The output item measure in the ministep provides detailed information regarding the analysis of these difficulty levels. The results of the item difficulty analysis are presented in Figure 2.



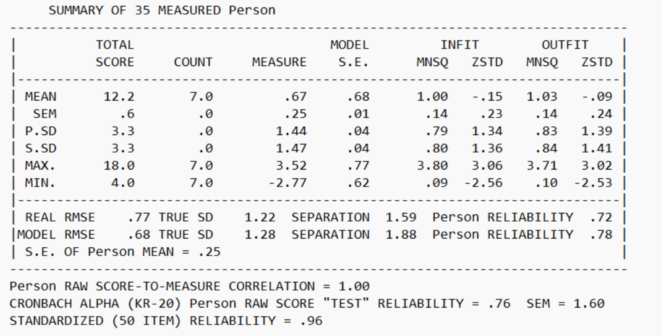
```
          Item STATISTICS:  MEASURE ORDER

-------------------------------------------------
|ENTRY   TOTAL  TOTAL     JMLE   MODEL|   INFIT  |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|
|------------------------------------+----------+
|     1     49     35     1.09    .29| .85  -.62|
|     2     57     35      .40    .30|1.12   .58|
|     6     57     35      .40    .30| .87  -.50|
|     3     58     35      .31    .30| .86  -.54|
|     7     59     35      .23    .30| .90  -.37|
|     4     65     35     -.31    .30| .64 -1.64|
|     5     83     35    -2.11    .33|1.95  3.35|
|------------------------------------+----------+
| MEAN    61.1   35.0      .00    .30|1.03   .04|
| P.SD     9.9     .0      .94    .01| .40  1.48|
-------------------------------------------------
```

Figure 2. Item measure

As shown in Figure 2, the mean logit value is 0.00 with a standard deviation of 0.94. The obtained value of 0.94 indicates that logit values greater than 0.94 categorize items as 'very difficult,' logit values between 0.0 and 0.94 are categorized as 'difficult,' logit values between 0.0 and -0.94 represent 'easy' items, and logit values below -0.94 are categorized as 'very easy.' Based on the item measure analysis, it was found that Item 1 is categorized as very difficult. Items categorized as difficult include Items 2, 6, 3, and 7. The item categorized as easy is Item 4, while Item 5 is categorized as very easy.

Discrimination power

Discrimination power in Rasch modelling explains how effectively a test item distinguishes between high-ability and low-ability individuals [30]. The results of the discriminatory power analysis can be found in the Ministep application under the output table menu, specifically in the summary statistics section. The relevant data to observe is the separation value, as shown in Figure 3.



```
     SUMMARY OF 35 MEASURED Person

-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.       MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      12.2        7.0         .67    .68       1.00   -.15   1.03   -.09 |
|  SEM        .6         .0         .25    .01        .14    .23    .14    .24 |
| P.SD       3.3         .0        1.44    .04        .79   1.34    .83   1.39 |
| S.SD       3.3         .0        1.47    .04        .80   1.36    .84   1.41 |
| MAX.      18.0        7.0        3.52    .77       3.80   3.06   3.71   3.02 |
| MIN.       4.0        7.0       -2.77    .62        .09  -2.56    .10  -2.53 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .77 TRUE SD    1.22  SEPARATION  1.59  Person RELIABILITY  .72 |
|MODEL RMSE    .68 TRUE SD    1.28  SEPARATION  1.88  Person RELIABILITY  .78 |
| S.E. OF Person MEAN = .25                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .76  SEM = 1.60
STANDARDIZED (50 ITEM) RELIABILITY = .96
```

Figure 3. Summary statistic (separation)

The separation value in the Summary Statistics output can categorize the item's discriminatory power, where a higher separation value indicates better discriminatory power. The obtained person separation value is 1.59. The person separation index for the students is calculated as follows:

$$H = \frac{[(4 \times 1.59) + 1]}{3} = 2.45$$

With a person separation value of 2.45, the aforementioned equation yields a separation index of 2.50, which is rounded to 3. This indicates that the respondents can be categorized into three distinct groups: high-ability students, moderate-ability students, and low-ability students. The results of this categorization also demonstrate that the designed questions are capable of grouping students into three distinct levels.

Based on the results of the previous analysis, seven test items were declared valid through the empirical validity analysis. The reliability analysis yielded an item reliability of 0.88, which falls within the 'good' range. Furthermore, the item difficulty analysis indicates that one item is categorized as difficult, five items as moderate, and one item as very easy. The results for item discriminatory power show four distinct levels of discrimination, all of which are considered good according to the model.

Conclusion

This study successfully developed a chemical literacy instrument for high school students on atomic structure material using the Rasch model, incorporating content, context, HOTS, and affective aspects within a four-level literacy assessment rubric. The analysis results demonstrate that the seven developed essay items are both theoretically and empirically valid and highly reliable. Furthermore, the instrument exhibits effective item difficulty distribution and discriminatory power to categorize student abilities, confirming its feasibility as a tool for measuring chemical literacy in atomic structure. Based on the results of this study, it is suggested that chemistry teachers utilize this instrument as an alternative assessment to measure students' chemical literacy levels more comprehensively. Furthermore, future researchers are encouraged to expand the scope of this instrument to other chemical topics or to involve a larger, more diverse sample of students to further strengthen its empirical evidence.

Author's Contribution

S. Faizah: developed the research design, conducted the research, analyzed the data, and finalized the writing of the scientific article. Yerimadesi: provided constructive suggestions to improve the writing. N. Yuhelman & Iryani: validator of chemistry literacy instruments.

Acknowledgements

The authors would like to express their gratitude to the Direktorat Penelitian dan Pengabdian kepada Masyarakat Direktorat Jenderal Riset dan Pengembangan Kementerian Pendidikan Tinggi, Sains dan Teknologi, for funding this research under contract number 088/C3/DT.05.00/PL/2025.

References

[1] S. Rahayu, “Prosiding Seminar Nasional Kimia UNY 2017 Sinergi Penelitian dan Pembelajaran untuk Mendukung Pengembangan Literasi Kimia pada Era Global Ruang Seminar FMIPA UNY, 14 Oktober,” 2017.

[2] S. Alfonso, M. Gesto, and B. Sadoul, “Temperature increase and its effects on fish stress physiology in the context of global warming,” *J. Fish Biol.*, vol. 98, no. 6, 2021, doi: 10.1111/jfb.14599.

[3] OECD, *PISA 2022 Results (Volume I)*. in PISA. OECD, 2023. doi: 10.1787/53f23881-en.

[4] OECD, “PISA 2022 Results Factsheets Indonesia PUBE,” 2023. [Online]. Available: https://oecdch.art/a40de1dbaf/C108.

[5] P. Permatasari and Z. Fitriza, “Analisis literasi sains siswa madrasah aliyah pada aspek konten, konteks, dan kompetensi materi larutan penyangga,” *EduKimia*, vol. 1, no. 1, pp. 53–9, Apr. 2019.

[6] N. Sutrisna, “Analisis Kemampuan Literasi Sains Peserta Didik SMA di Kota Sungai Penuh,” *Jurnal Inovasi Penelitian*, vol. 1, no. 12, p. 2683, 2021.

[7] R. A. Barus, A. Rusilowati, and S. Ridlo, “Analisis Kebutuhan Pengembangan Instrumen Tes Penilaian Literasi Sains Berorientasi TIMSS Siswa SD Kelas V,” *Jurnal Pemikiran dan Pengembangan Sekolah Dasar*, vol. 12, no. 1, pp. 68–85, 2024, doi: 10.22219/jp2sd.

[8] Y. Shwartz, R. Ben-Zv, and A. Hofstein, “The importance of involving high-school chemistry teachers in the process of defining the operational meaning of ‘chemical literacy,’” *Int. J. Sci. Educ.*, vol. 27, no. 3, pp. 323–344, 2005, doi: 10.1080/0950069042000266191.

[9] R. Chang, *Chemistry 10th Edition*. 2010.

[10] R. H. . Petrucci, F. Geoffrey. Herring, J. D. . Madura, and Carey. Bissonnette, *General chemistry : principles and modern applications*. Pearson, 2017.

[11] N. D. Jespersen, J. E. Brady, and A. Hylsop, *General Chemistry-the molecular nature of matter-6th ed*, 6th ed. New York, 2012.

[12] N. F. Diannisa, E. Erlina, A. I. Harun, R. Sahputra, and M. Ulfah, “Deskripsi Pemahaman Konsep pada Materi Struktur Atom di Kelas X SMA Negeri 01 Ngabang,” *Hydrogen: Jurnal Kependidikan Kimia*, vol. 11, no. 4, p. 497, Aug. 2023, doi: 10.33394/hjkk.v11i4.8091.

[13] J. Brady, *Chemistry The Molecular Nature of Matter (7th ed.)*, vol. 11, no. 1. 2014.

[14] H. La Muhamad, N. A. Limatahu, and M. E. Sriwahyuni, “Analisis Hasil Belajar Siswa Kelas X SMA Negeri 1 Sanana Pada Materi Struktur Atom,” *Jurnal Pendidikan Kimia Unkhair (JPKU)*, vol. 4, no. 2, pp. 10–20, Dec. 2024.

[15] N. Afrianis and L. Ningsih, “Analisis Kesulitan Belajar Siswa Pada Materi Struktur Atom,” *Jurnal Pendidikan Kimia dan Terapan*, vol. 6, no. 2, 2022.

[16] S. A. D. Putri and E. Yusmaita, “Pengembangan Instrumen Tes Literasi Kimia pada Materi Struktur Atom dan Nanoteknologi Fase E,” *Entalpi Pendidikan Kimia*, pp. 25–31, 2024.

[17] R. S. Anam, “Instrumen Penelitian yang Valid dan Reliabel,” *Jurnal Edukasi Sebelas April*, vol. 1, p. 1, Feb. 2017.

[18] F. A. Slamet, *Model Penelitian Pengembangan (RnD)*. Malang: Institut Agama Islam Sunan Kalijogo Malang, 2022.

[19] X. Liu, *Using and Developing Measurement Instruments in Science Education*, 2nd ed. 2020.

[20] F. R. Wilson, W. Pan, and D. A. Schumsky, “Recalculation of the critical values for Lawshe’s content validity ratio,” *Measurement and Evaluation in Counseling and Development*, vol. 45, no. 3, pp. 197–210, Jul. 2012, doi: 10.1177/0748175612440286.

[21] B. Sumintono and W. Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*, 1st ed. Trim Komunikata, 2015. [Online]. Available: www.trimkomunikata.com

[22] H. A. Saputri, Zulhijrah, nabila J. Larasati, and Shaleh, “Analisis Instrumen Assesmen: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Beda Butir Soal,” *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, vol. 9, no. 5, Dec. 2023.

[23] N. Eralita and Y. A. Setiawan, “Studi Literasi Kimia Siswa Sekolah Menengah Kejuruan pada Materi Zat dan Perubahan,” *Jurnal Eduscience (JES)*, vol. 9, no. 3, pp. 792–801, Dec. 2022.

[24] F. Yusup, “Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif,” *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, vol. 7, no. 1, pp. 17–23, 2018.

[25] F. Kurniawan, P. S. Pane, and R. Muhdi, “Analisis Soal Ulangan Harian Pelajaran Matematika Materi KPK dan FPB Siswa Kelas 5,” *Indonesian Journal of islamic Elementary Education*, vol. 4, no. 2, pp. 138–8, 2024.

[26] M. F. Ramadhan, R. A. Siroj, and M. W. Afgani, “Validitas and Reliabilitas,” *Journal on Education*, vol. 06, no. 02, pp. 10967–10975, 2024.

[27] Z. Arifin, “Kriteria Instrumen dalam suatu Penelitian,” *THEOREMS (The Original Research of Mathematics)*, vol. 2, no. 1, pp. 28–36, Jul. 2017.

[28] B. Sumintono, “Penilaian Keterampilan Berpikir Tingkat Tinggi: Aplikasi Pemodelan Rasch pada Asesmen Pendidikan,” Sep. 2016. [Online]. Available: http://deceng3.wordpress.com

[29] J. M. Linacre, “Optimizing Rating Scale Category Effectiveness,” 2002.

[30] E. D. K. Sari and I. Mahmudi, *Analisis Pemodelan Rasch Pada Assessment Pendidikan*. PT. Pena Persada Kerta Utama, 2024.