JPPIPA 11(7) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Systematic Literature Review: Three-Tier Diagnostic Test to Identifying Misconceptions in Chemistry

Beatrice Ruth Nathania Simanjuntak[1], Wiji Wiji[1*], Tuszie Widhiyanti[1]

[1]Chemistry Education, Universitas Pendidikan Indonesia, Bandung, Indonesia.

Received: March 30, 2025
Revised: May 07, 2025
Accepted: July 25, 2025
Published: July 31, 2025

Corresponding Author:
Wiji
maswiji@upi.edu

DOI: 10.29303/jppipa.v11i7.11011

Abstract: This study presents a literature review on the use of the three-tier diagnostic test in identifying students' misconceptions in chemistry and offers insights into its effectiveness in educational evaluation. A literature review of studies from 2019 to 2024 shows that 71.43% of these tests use close-ended questions due to their efficiency in terms of time, cost, and data analysis, compared to 28.57% that use open-ended questions. Research findings indicate that this test has successfully identified misconceptions in various chemistry topics, such as equilibrium, chemical bonding, ionization energy, and acids-bases. The ability of the three-tier diagnostic test to detect misconceptions makes it an effective tool for improving students' conceptual understanding. The findings indicate that the three-tier diagnostic test has been widely used to uncover students' conceptual difficulties across various chemistry topics. Its use has shown promise in supporting more accurate and in-depth assessments compared to traditional evaluation methods. The review suggests that this type of diagnostic testing can enhance the quality of formative assessments and contribute to improved instructional planning.

Keywords: Chemistry; Close-ended; Misconception; Open-ended; Three-tier diagnostic test

Introduction

Misconceptions in chemistry refer to students' incorrect or alternative conceptions that deviate from scientifically accepted knowledge and are often resistant to change through conventional instruction (Barke et al., 2009; Winarni et al., 2022). Identifying these misconceptions is crucial, as they can hinder meaningful learning and the development of accurate scientific understanding.

In the context of chemistry learning, such misconceptions frequently emerge due to the complex nature of the subject, which involves the integration of three levels of representation: macroscopic, submicroscopic, and symbolic (Johnstone, 1993; Treagust et al., 2003). The macroscopic level involves observable phenomena, such as color changes in chemical reactions; the submicroscopic level describes invisible particles like atoms and molecules; while the symbolic level uses notations such as chemical equations to represent those processes (Wiji et al., 2021).

Students' difficulties in coordinating these three representational levels often hinder conceptual understanding and contribute to the development of significant misconceptions, especially in abstract chemistry topics (Wang et al., 2019; Mardiyyaningsih et al., 2023). Therefore, assessing and diagnosing these misconceptions becomes a critical step in improving chemistry instruction.

One effective tool that has gained recognition in this context is the three-tier diagnostic test. This test incorporates three components: the student's answer, the reasoning behind it, and the level of confidence in their response (Caleon & Subramaniam, 2010; Gurel et

How to Cite:
Simanjuntak, B. R. N., Wiji, W., & Widhiyanti, T. (2025). Systematic Literature Review: Three-Tier Diagnostic Test to Identifying Misconceptions in Chemistry. *Jurnal Penelitian Pendidikan IPA*, *11*(7), 37–47. https://doi.org/10.29303/jppipa.v11i7.11011

al., 2015; Sari et al., 2024). By including a confidence tier, this test provides a more comprehensive and accurate diagnostic approach than the conventional two-tier format (Rismaningsih & Nurhafsari, 2022). It allows educators to differentiate between correct answers given with certainty and those based on guesses, thereby reducing the chance of misinterpreting students' actual understanding (Yusrizal & Halim, 2017; Meiliyadi et al., 2023).

This study offers a literature-based analysis not only of the use of three-tier diagnostic tests in chemistry education but also introduces a novel perspective by investigating how misconceptions are classified, how different question formats influence student responses, and what validation methods are employed to ensure the tests' accuracy and reliability. This expanded focus provides critical insights for educators and researchers aiming to develop more effective assessment strategies in science education.

Method

A literature review is an integral part of academic research aimed at providing a comprehensive overview of existing literature related to a specific theme, theory, or method. It strengthens the foundation of knowledge by synthesizing previous studies and identifying gaps in existing research (Paul & Criado, 2020; Lim et al., 2022). The stages of conducting a literature review include:

Preparation and Planning

This involves defining the objectives and scope of the review, identifying the topic of study, and determining the research questions to be answered (Cooper et al., 2018; Linnenluecke et al., 2020). Subsequently, a systematic search for relevant literature is conducted, utilizing appropriate search strategies and selecting relevant databases (Cooper et al., 2018).

The initial screening process was conducted by reviewing the titles and abstracts of the identified articles to assess their relevance to the research objectives. Articles that met the inclusion criteria based on title and abstract were then subjected to full-text screening. This step involved a more detailed evaluation to ensure alignment with the inclusion and exclusion criteria. Two independent reviewers conducted the screening process to minimize bias. In case of disagreement regarding article eligibility, the reviewers discussed until consensus was reached. The article criteria are shown in Table 1.

Table 1. Criteria for articles used in SLR

Criteria	Description
Included	Articles addressing three-tier diagnostic tests in chemistry.
	Articles published between 2019-2024.
	Articles focus on students' understanding rather than pre-service chemistry teachers.
	Articles indexed in Sinta or Scopus.
	Opened access articles.
Excluded	Literature reviews, nomenclature articles, and conference proceedings.
	Articles not discussing three-tier diagnostic tests in chemistry.
	Articles that focus on the understanding of pre-service chemistry teachers.
	Articles not indexed in Sinta or Scopus.
	Closed access articles.

Conducting the Review

After collecting the literature, the next step is to analyze and synthesize the findings. This involves critically evaluating existing studies and integrating relevant findings (Wolfswinkel et al., 2013). Findings are organized based on specific themes or categories to facilitate understanding and presentation (Smith, 2018). Articles were searched using Publish or Perish and Google Scholar, indexed in Sinta or Scopus. Between 2019 and 2024, a total of 40 articles related to the three-tier diagnostic test were identified. When categorized based on the predefined criteria, 14 articles met the criteria, while the remaining 26 articles did not. The keywords used for the article search were three-tier diagnostic, misconception, and chemistry. The flowchart illustrating the implementation stages is presented in Figure 1.

Table 2. Articles used in the SLR

Article Title	Writer	Code	Year	Journal	Indexed
The Misconception Diagnosis on Ionic and Covalent Bonds Concepts with Three Tier Diagnostic Test	Anti Kolonial Prodjosantoso, Artanti Mulia Hertina and Irwanto	CE	2019	International Journal of Instruction	Scopus, Google Scholar
Analisis Pemahaman Konsep pada Pembelajaran Hidrolisis Berbantuan Metode Blended Learning Berbasis Inkuiri Terbimbing	Sri Winarsih and Sigit Priatmoko	OE	2019	Chemistry in Education	Sinta 5
Developing a three-tier diagnostic instrument on chemical equilibrium (TT-DICE)	J. Jusniar, E. Effendy, Endang Budiasih and S. Sutrisno	CE	2020	Educación Química	Scopus (Q3)

Article Title	Writer	Code	Year	Journal	Indexed
Using online three-tier diagnostic test to assess conceptions of ionization energy	Nadi Suprapto and Azmil Abidah	CE	2020	Periodico Tche Quimica	Scopus, Google Scholar
Misconception Analysis of Buffer Material using Three Tier Multiple Choice Test assisted by CBT for SMAN 9 Semarang	Ifandika Dwi Septian, Endang Susilaningsih and Sri Susilogati Sumarti	CE	2020	Journal of Innovative Science Education	Sinta 3
Analisis Miskonsepsi Peserta Didik pada Materi Asam Basa Menggunakan Instrumen Three-Tier Diagnostic Test	Ekawisudawati, Mohammad Wijaya and Muhammad Danial	CE	2021	Chemistry Education Review	Sinta 5
Analysis of Understanding Chemical Bond Concepts in Students with Three-Tier Multiple Choice	Mellyzar	CE	2021	Journal of Educational Chemistry	Sinta 3
Identification and Analysis of Students' Misconceptions Using Three-Tier Multiple Choice Diagnostic Instruments on Thermochemistry Topic	Irfandi, Rosa Murwindra, Dwi Putri Musdansi, Wimbi Apriwanda N, Chuzairy Hanri	OE	2022	IJECA (International Journal of Education & Curriculum Application)	Sinta 3
Identification of Misconceptions in Chemical Bonding Materials Using Three Tier Diagnostic Test	Nur Candra Eka Setiawan and Putri Ridha Ilahi	CE	2022	Journal of Natural Science and Integration	Sinta 2
Development of Three Tier Open-Ended Instrument to Measure Chemistry Students' Critical Thinking Disposition Using Rasch Analysis	Wahyudi, A., Richardo, R., Eilks, I., & Kulgemeyer, C	OE	2023	International Journal of Instruction	Scopus, Google Scholar
Development of Three-Tier Diagnostic Test Instrument to Measure Misconceptions of Class XI Students on Reaction Rate Materials	Putri Dewi Natalia and Ajat Sudrajat	CE	2023	Jurnal Teknologi Pendidikan	Sinta 4
Identification of Student Misconceptions Using a Three-tier Test on the Concept of Atoms, Ions, and Molecules	Rieke Oktavia Fanfiana, Saprizal Hadisaputra and Supriadi	CE	2023	Chemistry Education Practice	Sinta 4
Analysis of Understanding the Concept of Alkenes through the Three-tier Multiple Choice Diagnostic Test Instrument	Selina, Rini Muharini, Ira Lestari, Masriani and Rahmat Rasmawan	CE	2024	Jurnal Penelitian Pendidikan IPA	Sinta 2
Development and Validation of a Three-Tier Test for Identifying Misconceptions in Organic Chemistry Course	Asyti Febliza, Asep Kadarohman, Siti Aisyah, Norazilawati Abdullah	OE	2024	Journal of Innovative Science Education	Sinta 3

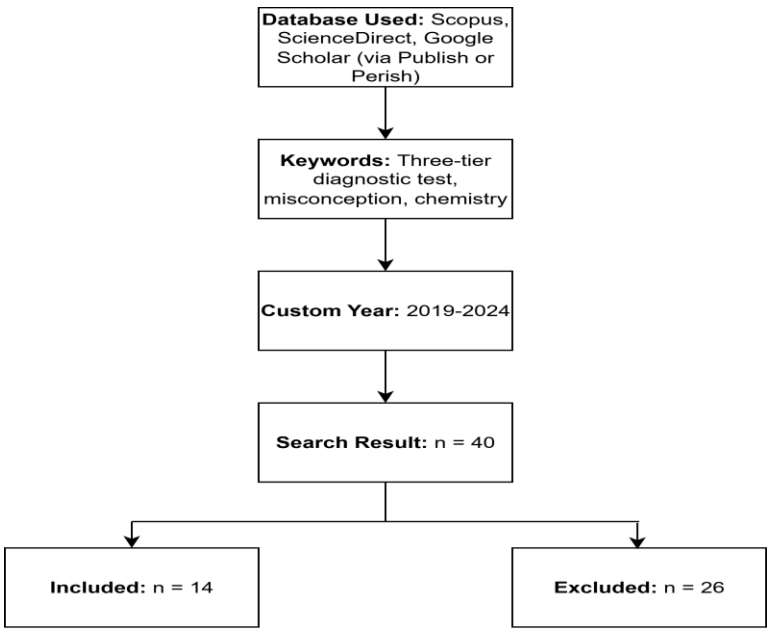

Figure 1. SLR implementation stages

Writing and Reporting

Writing the literature review involves presenting a summary of the analyzed literature, including the methodologies used and key findings (Cooper et al., 2018; Linnenluecke et al., 2020). At this stage, the results of the SLR are reported based on the previously formulated research questions. Articles categorized as included are labelled as CE (close-ended) and OE (open-ended). The list of articles used in this SLR is presented in Table 2.

Result and Discussion

Patterns of Close-Ended and Open-Ended Questions in the Three-Tier Diagnostic Test

A literature review on the three-tier mental model diagnostic test provides significant insights into the effectiveness of this instrument in identifying misconceptions in complex chemistry concepts. Based on the analysis of four articles that met the selection criteria (Table 2), several key findings are presented in Table 3.

Table 3. Patterns of close-ended and open-ended questions in the three-tier diagnostic test

Patterns	N	Percentage (%)
Open Ended	4	28.57
Close Ended	10	71.43
Total	14	100

The articles reviewed in this study were classified based on the type of questions used in the three-tier diagnostic test instrument: open-ended (OE) and close-ended (CE). Out of 14 analyzed articles, 10 articles (71.43%) used close-ended questions, while 4 articles (28.57%) used open-ended questions. These results indicate that most studies prefer the close-ended approach in the three-tier diagnostic test.

The difference between open-ended and close-ended questions lies in their format. In the close-ended format, all available answers are in multiple-choice form or predetermined. In other words, students select from given options, both for the main answer and the supporting reason. In the open-ended format, students have the freedom to respond in descriptive or short-essay form, particularly in tier 2 (reasoning) and tier 3 (confidence level).

The advantages of open-ended questions are: allowing students to express their understanding more deeply and provide more diverse responses (Birenbaum & Tatsuoka, 1987; Bingölbali et al., 2021) students can explain their thoughts freely (Desai & Reimers, 2019; Bingölbali et al., 2021) open-ended questions can be more effective in identifying students' conceptual errors (Gurel et al., 2015).

The advantages of close-ended questions are: more efficient in terms of time and cost; facilitating data analysis because responses are already structured and can be directly processed statistically (Baburajan et al., 2022; Desai & Reimers, 2019); close-ended questions can provide more consistent and easier results compared to open-ended questions (Baburajan et al., 2022).

The disadvantages of open-ended questions are: requiring more time and resources for analysis because responses must be coded and analyzed manually (Baburajan et al., 2022); highly varied responses can make analysis more complex and challenging (Bingölbali et al., 2021).

The disadvantages of close-ended questions are: they may limit students' responses, which may not fully reflect their understanding (Desai & Reimers, 2019); they are not as effective as open-ended questions in exploring deep understanding and conceptual errors (Birenbaum & Tatsuoka, 1987).

Both open-ended and close-ended questions in the three-tier diagnostic test have their own strengths and weaknesses. Open-ended questions are better suited for exploring deep understanding and identifying conceptual errors, whereas close-ended questions are more efficient and easier to analyze. The choice of format depends on the specific objectives of the diagnostic test being conducted.

Validity and Reliability of the Three-Tier Diagnostic Test in Various Chemistry Topics

The three-tier diagnostic test is an evaluation tool used to identify students' misconceptions in various chemistry concepts. The reliability and accuracy depend on the validity and reliability of the instrument used. Based on an analysis of various articles, several methods for testing validity and reliability were identified, as presented in Table 4.

The differences in validity methods across studies reflect adjustments to measurement goals, sample size, and available resources. The use of Content Validity Ratio (CVR) and Confirmatory Factor Analysis (CFA) represents a quantitative approach suited for psychological constructs like critical thinking disposition (Wahyudi et al., 2023), while simpler instruments often rely on validation by two experts (Prodjosantoso et al., 2019).

CFA is particularly appropriate when the sample size is large (typically around 250–500 respondents) and randomly selected, as the method is highly sensitive to sample size and is used to examine whether items align with a hypothesized construct structure. This makes CFA ideal for validating latent psychological variables in a measurement model (Swarni et al., 2024; Gusmanida et al., 2024).

Table 4. Validity and reliability used in several articles

Article Title	Author	Validity	Reliability
The Misconception Diagnosis on Ionic and Covalent Bonds Concepts with Three Tier Diagnostic Test.	Prodjosantoso et al. (2019)	Instrument validity was tested by two experts to ensure concept accuracy and instrument appropriateness	Not mentioned
Developing a three-tier diagnostic instrument on chemical equilibrium (TT-DICE)	Jusniar et al. (2020)	Content validity and item validity. Content validation was conducted by six expert validators to assess instrument suitability with concept indicators, language clarity, and concept accuracy. Item validity was tested using Pearson correlation	Reliability test was conducted using Cronbach's Alpha, measuring the internal consistency of each tier in TT-DICE
Using online three-tier diagnostic test to assess conceptions of ionization energy	Suprapto (2020)	Content validity was conducted by adapting instruments from previous studies (Suprapto et al., 2018; Tan et al., 2005; Taber & Tan, 2011) and translating them into Indonesian	Instrument reliability was tested by observing the decrease in misconceptions from the first tier to the third tier in the Three-Tier Diagnostic Test
Development of Three Tier Open-Ended Instrument to Measure Chemistry Students' Critical Thinking Disposition Using Rasch Analysis	Wahyudi et al. (2023)	Construct validity was analyzed using Confirmatory Factor Analysis (CFA), and content validity was tested using the Content Validity Ratio (CVR)	Instrument reliability was tested using Rasch Analysis
Analisis Pemahaman Konsep pada Pembelajaran Hidrolisis Berbantuan Metode Blended Learning Berbasis Inkuiri Terbimbing	Winarsih & Priatmoko (2019)	Conducted but not explained in detail	Conducted but not explained in detail.
Misconception Analysis of Buffer Material using Three Tier Multiple Choice Test assisted by CBT for SMAN 9 Semarang	Septian et al. (2020)	Content and item validation were conducted with two lecturers and one teacher from SMAN 9 Semarang	Reliability test was calculated using the KR-21 formula
Analisis Miskonsepsi Peserta Didik pada Materi Asam Basa Menggunakan Instrumen Three-Tier Diagnostic Test	Ekawisudawati et al. (2021)	Conducted but not explained in detail	Conducted but not explained in detail
Analysis of Understanding Chemical Bond Concepts in Students with Three-Tier Multiple Choice	Mellyzar (2021)	Validated by five expert lecturers. The validators were lecturers teaching General Chemistry, Chemical Bonding, and Inorganic Chemistry	Not explained
Identification and Analysis of Students' Misconceptions Using Three-Tier Multiple Choice Diagnostic Instruments on Thermochemistry Topic	Irfandi et al. (2022)	Not explained	Not explained
Identification of Misconceptions in Chemical Bonding Materials Using Three Tier Diagnostic Test	Setiawan & Ridha (2022)	The three-tier test instrument used was developed by Mutiara Ismet (2015) and has undergone validity, reliability, discrimination power, and difficulty level testing	The instrument was adopted from previous research and had already been tested for reliability
Development of Three-Tier Diagnostic Test Instrument to Measure Misconceptions of Class XI Students on Reaction Rate Materials	Natalia & Sudrajat (2023)	Validation was conducted by expert validators whose number was not specified. Validity was tested using the biserial coefficient to calculate the correlation between item scores and the total test score	Instrument reliability was calculated using the Kuder-Richardson (K-R-20) formula

Article Title	Author	Validity	Reliability
Identification of Student Misconceptions Using a Three-tier Test on the Concept of Atoms, Ions, and Molecules	Fanfiana et al. (2024)	Content validity was validated by five experts: two chemistry lecturers from UIN Sunan Kalijaga and Universitas Negeri Yogyakarta, along with five middle school science teachers in Yogyakarta. Construct Validity: Validity testing was conducted using the Product-Moment correlation with statistical tests on 30 twelfth-grade students at MA Al-Ishlahuddiny Putra Kediri	Conducted but not explained in detail
Analysis of Understanding the Concept of Alkenes through the Three-tier Multiple Choice Diagnostic Test Instrument	Selina et al. (2024)	Validity of language, material, and interview tests was validated by three experts. The validation results were calculated using Aiken's formula with the help of Microsoft Excel	Reliability test was conducted on 20 Chemistry Education students at Tanjungpura University (Untan) using the KR-20 formula
Development and Validation of a Three-Tier Test for Identifying Misconceptions in Organic Chemistry Course	Febliza et al. (2024)	Content validity was tested using the Content Validity Index (CVI) by three experts, and empirical validity was tested using the Pearson Product-Moment validity test on 25 students	Reliability was tested using Cronbach's Alpha

In contrast, CVR offers a quantitative measure of content relevance through expert consensus but lacks empirical validation of item performance. Meanwhile, Pearson correlation provides empirical validity by assessing the relationship between item scores and total scores, yet it does not confirm the conceptual appropriateness of each item.

The Rasch model, on the other hand, simultaneously evaluates item difficulty and student ability, making it more suitable for performance-based instruments such as diagnostic tests. Pearson correlation, as applied by Jusniar et al. (2020), is often used in three-tier multiple-choice tests to identify student misconceptions, particularly when working with numerical data that follow a linear and normal distribution (Santoso et al., 2017; Mulyana & Desnita, 2023).

Based on the findings of a study that employed CVR and CFA for validity and Rasch analysis for reliability, these three methods provided complementary contributions in the process of instrument development and validation (Wahyudi et al., 2023). The use of the Content Validity Ratio (CVR) in the early development stage enabled researchers to ensure that the items aligned with the predetermined construct indicators through expert judgment (Farzad et al., 2020). Subsequently, Confirmatory Factor Analysis (CFA) was applied to assess model fit and confirm that each item had a significant relationship with the latent construct being measured, thereby supporting the construct validity of the instrument (Marsh et al., 2020; Vaingankar et al., 2020).

However, considering the limitations of CFA in comparing groups or generalizing across populations (Lu & Bi, 2016), the Rasch analysis was used as a complementary approach. As part of Probabilistic Test Theory, Rasch analysis allows for the simultaneous evaluation of item and person fit, while also providing additional information on item difficulty, unidimensionality, and instrument reliability (Wren & Barbera, 2014; Hale et al., 2016). Thus, the combination of CVR, CFA, and Rasch ensures that the developed instrument is not only content valid but also empirically sound and consistent in measurement (Behmke & Atwood, 2013; Farzad et al., 2020).

The KR-20 (Kuder-Richardson Formula 20) is a method used to measure the reliability of instruments consisting of multiple-choice items with only one correct answer (Fanani et al., 2023; Selina et al., 2024). It is suitable when the difficulty level of each item is not guaranteed to be equal. KR-20 is often used in educational research to ensure internal consistency of tests.

Meanwhile, the approach of adapting existing instruments (Suprapto, 2020) relies more heavily on historical validity or dependence on previous studies, which can expedite development but poses risks if the cultural and linguistic contexts are not critically adapted.

Several studies, claim that validity and reliability tests were conducted but do not provide detailed methods or results (Ekawisudawati et al., 2021; Winarsih & Priatmoko, 2019). Such omissions reduce scientific transparency and hinder readers from evaluating the trustworthiness and dependability of the instrument. Without these details, confidence in the findings is limited, and the potential for replication by other researchers remains low.

Identifying Students' Misconceptions Using the Three-Tier Diagnostic Test

The three-tier test allows educators to distinguish between students who hold strong confidence in incorrect concepts (misconceptions) and those who are unsure and lack sufficient understanding (lack of knowledge), by analyzing the combination of answers, reasoning, and confidence levels (Istiyani et al., 2018; Prodjosantoso et al., 2019; Shiddiqi et al., 2024). The categories of students' conceptual understanding categories in the three-tier diagnostic test are explained in Table 5.

Table 5. Students' conceptual categories in the three-tier diagnostic test

Category	Response Type	Code
Scientific Knowledge	Correct response + scientific reasoning + confident	SK
Lack of Knowledge	Correct response + scientific reasoning + not confident	LoK
	Incorrect response + scientific reasoning + not confident	
	Correct response + non-scientific reasoning + not confident	
	Incorrect response + non-scientific reasoning + not confident	
Error	Incorrect response + scientific reasoning + confident	E
Misconception	Correct response + non-scientific reasoning + confident	M
	Incorrect response + non-scientific reasoning + confident	

Chemistry topics and misconceptions identified using the three-tier diagnostic test from various studies are shown in Table 6.

Table 6. Students' misconceptions in various chemistry topics

Chemistry Topic	Misconceptions	Source
Ionic and Covalent Bonds	Ionization concept, properties of ionic bonds, strength of covalent bonds, boiling points, and bond formation	Prodjosantoso et al. (2019)
Chemical Equilibrium	Dynamic equilibrium, effect of temperature, effect of pressure, and effect of concentration	Jusniar et al. (2020)
Ionization Energy	Energy conservation, half-filled or fully filled sub-shell stability, and the octet rule	Suprapto (2020)
Acid-Base	Bronsted-Lowry acid-base theory, pH concept, acid-base reactions, and the strength of acids and bases	Wahyudi et al. (2023), Ekawisudawati et al. (2021)
Hydrolysis	Role of hydrolysis in daily life, determining hydrolysis reactions, identifying types of hydrolysis, and calculating pH	Winarsih & Priatmoko (2019)
Buffer Solution	Buffer concept, effect of concentration and volume, examples of buffers in daily life, and calculating buffer pH	Septian et al. (2020)
Chemical Bonds	Basic concepts of chemical bonding, structure and properties of bonds, process of forming ionic and covalent bonds, and the relationship between metallic bonds and the physical properties of metals.	Mellyzar (2021) and Setiawan & Ridha (2022)
Thermochemistry	System and surroundings, chemical equilibrium, and general thermochemistry concepts	Irfandi et al. (2022)
Reaction rate	Collision theory, reaction order, and reaction rate equations	Natalia & Sudrajat (2023)
Atoms, Ions, and Molecules	Atomic structure and properties, electron transfer processes, and differences between molecular compounds and elements	Fanfiana et al. (2024)
Alkenes	Alkenes in everyday life, molecular formula of alkenes, alkene structures, alkene nomenclature based on IUPAC rules, physical and chemical properties of alkenes, and alkene formation reactions	Selina et al. (2024)
Organic Chemistry	Naming of organic compounds, physical and chemical properties of organic compounds, reaction mechanisms of SN1, SN2, E1, and E2	Febliza et al. (2024)

The analysis of students' response patterns across the three tiers can reveal specific misconceptions in various chemistry concepts. For example, in chemical equilibrium, students who misunderstand dynamic equilibrium tend to provide incorrect answers with incorrect reasoning while remaining confident in their responses (Jusniar et al., 2020). Various studies have shown that three-tier diagnostic tests have successfully identified misconceptions in different chemistry topics, such as ionic and covalent bonds, ionization energy, acids and bases, hydrolysis, buffer solutions, and organic chemistry. Therefore, this diagnostic test can serve as an effective tool for identifying students' misconceptions, providing valuable insights that can inform subsequent instructional interventions to enhance conceptual understanding."

Conclusion

This systematic literature review shows that the three-tier diagnostic test has been widely applied to identify students' misconceptions across various chemistry topics, especially using close-ended formats due to their efficiency. The structure of the test, which includes answer selection, reasoning, and confidence level, enables more precise differentiation between misconceptions and a lack of knowledge. While it effectively reveals misconceptions, it should be noted that diagnosis alone does not directly address them; instructional follow-up is required. Therefore, the three-tier diagnostic test is a valuable diagnostic tool to inform targeted pedagogical interventions. Future research should emphasize the development of standardized guidelines to improve the quality, validity, and reliability of such instruments.

Acknowledgments
The researcher expressed his gratitude to all parties who have supported the implementation of this research, especially to Prof. Dr. Wiji, M.Si and Tuszie Widhiyanti, M.Pd., Ph.D who have provided support, insight, facilities, and opportunities in completing this article.

Author Contributions
Conceptualization, methodology, resources, writing—review and editing, B.R.N.S., W.W., and T.W.; formal analysis, investigation, writing—original draft preparation, visualization and project administration, B.R.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

References

Baburajan, V., de Abreu e Silva, J., & Pereira, F. C. (2022). Open vs Closed-Ended Questions in Attitudinal Surveys – Comparing, Combining, and Interpreting Using Natural Language Processing. *Transportation Research Part C: Emerging Technologies*, *137*, 103589. https://doi.org/10.1016/j.trc.2022.103589

Barke, H.-D., Hazari, A., & Yitbarek, S. (2009). Students' Misconceptions and How to Overcome Them. *Misconceptions in Chemistry*, 21–36. https://doi.org/10.1007/978-3-540-70989-3_3

Behmke, D. A., & Atwood, C. H. (2013). Implementation and Assessment of Cognitive Load Theory (CLT) Based Questions in an Electronic Homework and Testing System. *Chemistry Education Research and Practice*, *14*(3), 247–256. https://doi.org/10.1039/C3RP20153H

Bingölbali, E., Bingölbali, F., & Prof, A. (2021). An Examination of Open-Ended Mathematics Questions' Affordances. *International Journal of Progressive Education*, *17*(4), 1–16. https://doi.org/10.29329/ijpe.2021.366.1

Birenbaum, M., & Tatsuoka, K. K. (1987). Open-Ended Versus Multiple-Choice Response Formats—It Does Make a Difference for Diagnostic Purposes. *Applied Psychological Measurement*, *11*(4), 385–395. https://doi.org/10.1177/014662168701100404

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Cooper, C., Booth, A., Varley-Campbell, J., Britten, N., & Garside, R. (2018). Defining the Process to Literature Searching in Systematic Reviews: A Literature Review of Guidance and Supporting Studies. *BMC Medical Research Methodology*, *18*(1), 1–14. https://doi.org/10.1186/S12874-018-0545-3

Desai, S., & Reimers, S. (2019). Comparing the Use of Open and Closed Questions for Web-Based Measures of the Continued-Influence Effect. *Behavior Research Methods*, *51*(3), 1426–1440. https://doi.org/10.3758/s13428-018-1066-z

Ekawisudawati, E., Wijaya, M., & Danial, M. (2021). Analisis Miskonsepsi Peserta Didik pada Materi Asam Basa Menggunakan Instrumen Three-Tier Diagnostic Test. *Chemistry Education Review (CER)*, *5*(1), 62–72. https://doi.org/10.26858/cer.v5i1.26359

Fanani, R. D., Supardi, Z. A. I., & Suprapto, N. (2023). The Development of HOTS-Based Question Instrument in Temperature and Heat Material.

Jurnal Penelitian Pendidikan IPA, *9*(9), 6890–6895. https://doi.org/10.29303/jppipa.v9i9.3282

Fanfiana, R. O., Hadisaputra, S., & Supriadi, S. (2024). Identification of Student Misconceptions Using A Three-Tier Test on The Concept of Atoms, Ions, and Molecules. *Chemistry Education Practice*, *7*(1), 75–81. https://doi.org/10.29303/cep.v7i1.6133

Farzad, M., MacDermid, J. C., Lu, Z., & Shafiee, E. (2020). Validation of Persian Version of Patient-Rated Wrist and Hand Evaluation: Confirmatory Factor Analysis and Rasch Analysis. *Archives of Rehabilitation Research and Clinical Translation*, *2*(4), 100076. https://doi.org/10.1016/j.arrct.2020.100076

Febliza, A., Kadarohman, A., Aisyah, S., & Abdullah, N. (2024). Development and Validation of a Three-Tier Test for Identifying Misconceptions in Organic Chemistry Course. *Journal of Innovative Science Education*, *13*(3), 124–133. https://doi.org/10.15294/jise.v13i3.13578

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/ eurasia.2015.1369a

Gusmanida, G., Sujati, H., & Herwin, H. (2024). Testing the Construct Validity and Reliability of the Student Learning Motivation Scale Using Confirmatory Factor Analysis (CFA). *Jurnal Penelitian Pendidikan IPA*, *10*(7), 4227–4234. https://doi.org/10.29303/jppipa.v10i7.7518

Hale, L. V. A., Lutter, J. C., & Shultz, G. V. (2016). The Development of a Tool for Measuring Graduate Students' Topic Specific Pedagogical Content Knowledge of Thin Layer Chromatography. *Chemistry Education Research and Practice*, *17*(4), 700–710. https://doi.org/10.1039/c5rp00190k

Irfandi, I., Murwindra, R., Musdansi, D. P., N, W. A., & Hanri, C. (2022). Identification and Analysis of Students' Misconceptions Using Three-Tier Multiple Choice Diagnostic Instruments on Thermochemistry Topic. *IJECA (International Journal of Education and Curriculum Application)*, *5*(3), 306–316. https://doi.org/10.31764/ ijeca.v5I3.11613

Istiyani, R., Muchyidin, A., & Rahardjo, D. H. (2018). Analysis of Student Misconception on Geometry Concepts Using Three-Tier Diagnostic Test. *Cakrawala Pendidikan*, *37*(2), 261628. https://doi.org/10.21831/cp.v37i2.14493

Johnstone, A. H. (1993). The Development of Chemistry Teaching: A Changing Response to Changing Demand. *Journal of Chemical Education*, *70*(9), 701–705. https://doi.org/10.1021/ ed070p701

Jusniar, J., Effendy, E., Budiasih, E., & Sutrisno, S. (2020). Developing a Three-Tier Diagnostic Instrument on Chemical Equilibrium (TT-DICE). *Educación Química*, *31*(3), 84–102. https://doi.org/10.22201/ fq.18708404e.2020.3.72133

Lim, W. M., Kumar, S., & Ali, F. (2022). Advancing Knowledge Through Literature Reviews: 'What', 'Why', and 'How to Contribute.' *The Service Industries Journal*, *42*(7–8), 481–513. https://doi.org/10.1080/02642069.2022.2047941

Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Conducting Systematic Literature Reviews and Bibliometric Analyses. *Australian Journal of Management*, *45*(2), 175–194. https://doi.org/10.1177/0312896219877678

Lu, S., & Bi, H. (2016). Development of a Measurement Instrument to Assess Students' Electrolyte Conceptual Understanding. *Chemistry Education Research and Practice*, *17*(4), 1030–1040. https://doi.org/10.1039/c6rp00137h

Mardiyyaningsih, A. N., Erlina, E., Ulfah, M., & Wafiq, A. F. (2023). Validity and Reliability of the Two-tier Diagnostic Test to Identify Students' Alternative Conceptions of Intermolecular Forces. *Jurnal Penelitian Pendidikan IPA*, *9*(6), 4375–4381. https://doi.org/10.29303/jppipa.v9i6.2797

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal Balance Between Goodness of Fit and Parsimony. *Multivariate Behavioral Research*, *55*(1), 102–119. https://doi.org/10.1080/00273171.2019.1602503

Meiliyadi, L. A. D., Asyari, A., & Arizona, K. (2023). Identification of Tadris Biology Students Level Understanding and Misconceptions on the Material of Quantities and Units Using 3-Tier Diagnostic Method. *Jurnal Penelitian Pendidikan IPA*, *9*(12), 12042–12048. https://doi.org/10.29303/jppipa.v9i12.6122

Mellyzar, M. (2021). Analysis of Understanding Chemical Bond Concepts in Students with Three-Tier Multiple Choice. *JEC*, *3*(1), 53–66. https://doi.org/10.21580/ jec.2021.3.1.7560

Mulyana, V., & Desnita, D. (2023). Empirical Validity and Reliability of the Scientific Literacy Assessment Instrument Based on the Tornado Physics Enrichment Book. *Jurnal Penelitian Pendidikan IPA*, *9*(5), 3961–3967. https://doi.org/10.29303/jppipa.v9i5.3290

Natalia, P. D., & Sudrajat, A. (2023). Development of Three-Tier Diagnostic Test Instrument to Measure Misconceptions of Class XI Students on Reaction Rate Materials. *Jurnal Teknologi Pendidikan: Jurnal*

Penelitian dan Pengembangan Pembelajaran, *8*(1), 1–11. https://doi.org/10.33394/jtp.v8i1.6192

Paul, J., & Criado, A. R. (2020). The Art of Writing Literature Review: What Do We Know and What Do We Need to Know? *International Business Review*, *29*(4), 101717. https://doi.org/10.1016/j.ibusrev.2020.101717

Prodjosantoso, A. K., Hertina, A. M., & Irwanto, I. (2019). The Misconception Diagnosis on Ionic and Covalent Bonds Concepts with Three Tier Diagnostic Test. *International Journal of Instruction*, *12*(1), 1477–1488. https://doi.org/10.29333/iji.2019.12194a

Rismaningsih, F., & Nurhafsari, A. (2022). Identify Hydrostatic Misconceptions Using Four Tier Diagnostic Tests with the Help of iSpring Suite 9 (Case Study in UNIS Faculty of Engineering Students). *Jurnal Penelitian Pendidikan IPA*, *8*(2), 773–780. https://doi.org/10.29303/jppipa.v8i2.1290

Santoso, B., Marchira, C. R., & Sumarni, P. (2017). Development and Validity and Reliability Tests of Professionalism Assessment Instrument in Psychiatry Residents. *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, *6*(1), 60–65. https://doi.org/10.22146/jpki.25369

Sari, D. N., Arif, K., Yurnetti, Y., & Putri, A. N. (2024). Identification of Students' Misconceptions in Junior High Schools Accredited A Using the Three Tier Test Instrument in Science Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(1), 1–11. https://doi.org/10.29303/jppipa.v10i1.5064

Selina, S., Muharini, R., Lestari, I., Masriani, M., & Rasmawan, R. (2024). Analysis of Understanding the Concept of Alkenes through the Three-tier Multiple Choice Diagnostic Test Instrument. *Jurnal Penelitian Pendidikan IPA*, *10*(6), 3463–3472. https://doi.org/10.29303/jppipa.v10i6.6510

Septian, I. D., Susilaningsih, E., & Sumarti, S. S. (2020). A Misconception Analysis of Buffer Material Using Three Tier Multiple Choice Test assisted by CBT for SMAN 9 Semarang. *Journal of Innovative Science Education*, *9*(1), 40–49. https://doi.org/10.15294/jise.v8i1.32027

Setiawan, N. C. E., & Ilahi, P. R. (2022). Identification of Misconceptions in Chemical Bonding Materials Using Three Tier Diagnostic Test. *Journal of Natural Science and Integration*, *5*(1), 77–89. https://doi.org/10.24014/jnsi.v5i1.16860

Shiddiqi, M. H. A., Arthamena, V. D., Ayyubi, M., Manarisip, A. J., & Aznam, N. (2024). Systematic Literature Review: Analysis of Misconception Problems and Diagnostic Instruments for Learning Chemistry. *Jurnal Penelitian Pendidikan IPA*, *10*(4), 168–179. https://doi.org/10.29303/jppipa.v10i4.5189

Smith, B. (2018). Doing a Literature Review: Releasing the Research Imagination. *Journal of Perioperative Practice*, *28*(12), 318–318. https://doi.org/10.1177/1750458918810149

Suprapto, N. (2020). Do We Experience Misconceptions?: An Ontological Review of Misconceptions in Science. *Studies in Philosophy of Science and Education*, *1*(2), 50–55. https://doi.org/10.46627/sipose.v1i2.24

Swarni, A., Herwin, H., & Sujati, S. (2024). Testing the Construct Validity and Reliability of the Student Learning Interest Scale Using Confirmatory Factor Analysis (CFA). *Jurnal Penelitian Pendidikan IPA*, *10*(9), 6322–6330. https://doi.org/10.29303/jppipa.v10i9.8794

Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2003). The Role of Submicroscopic and Symbolic Representations in Chemical Explanations. *International Journal of Science Education*, *25*(11), 1353–1368. https://doi.org/10.1080/0950069032000070306

Vaingankar, J. A., Abdin, E., Dam, R. M. V., Chong, S. A., Tan, L. W. L., Sambasivam, R., Seow, E., Chua, B. Y., Wee, H. L., Lim, W. Y., & Subramaniam, M. (2020). Development and Validation of the Rapid Positive Mental Health Instrument (R-PMHI) for Measuring Mental Health Outcomes in the Population. *BMC Public Health*, *20*(1), 1–12. https://doi.org/10.1186/s12889-020-08569-w

Wahyudi, A., Richardo, R., Eilks, I., & Kulgemeyer, C. (2023). Development of Three Tier Open-Ended Instrument to Measure Chemistry Students' Critical Thinking Disposition Using Rasch Analysis. *International Journal of Instruction*, *16*(3), 191–204. Retrieved from https://e-iji.net/ats/index.php/pub/article/view/76

Wang, T. L., Dai, F., Hu, R., Seager, S., Febriyanti, F., Wiji, W., & Widhiyanti, T. (2019). Thermochemistry Multiple Representation Analysis for Developing Intertextual Learning Strategy Based on Predict Observe Explain (POE). *Journal of Physics: Conference Series*, *1157*(4), 042042. https://doi.org/10.1088/1742-6596/1157/4/042042

Wiji, W., Widhiyanti, T., Delisma, D., & Mulyani, S. (2021). The Intertextuality Study of the Conception, Threshold Concept, and Troublesome Knowledge on Redox Reaction. *Journal of Engineering Science and Technology*, *16*(2), 1356–1369. Retrieved from https://jestec.taylors.edu.my/Vol%2016%20issue%202%20April%202021/16_2_33.pdf

Winarni, S., Effendy, E., Budiasih, E., Wonorahardjo, S., & Winarni, S. (2022). Constructing 'Concept Approval Strategy,' A Chemistry Learning Idea to Prevent Misconceptions. *Educación Química*, *33*(2),

159–180. https://doi.org/10.22201/fq.18708404e.2022.2.79841

Winarsih, S., & Priatmoko, S. (2019). Analisis Pemahaman Konsep Menggunakan Three-Tier Multiple Choice Test pada Pembelajaran Hidrolisis Berbantuan Metode Blended Learning Berbasis Inkuiri Terbimbing. *Chemistry in Education*, *8*(2), 29–36. Retrieved from https://journal.unnes.ac.id/sju/chemined/article/view/39128

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using Grounded Theory as a Method for Rigorously Reviewing Literature. *European Journal of Information Systems*, *22*(1), 45–55. https://doi.org/10.1057/ejis.2011.51

Wren, D., & Barbera, J. (2014). Psychometric Analysis of the Thermochemistry Concept Inventory. *Chemistry Education Research and Practice*, *15*(3), 380–390. https://doi.org/10.1039/c3rp00170a

Yusrizal, Y., & Halim, A. (2017). The Effect of The One-Tier, Two-Tier, and Three-Tier Diagnostic Test Toward the Students' Confidence and Understanding Toward the Concepts of Atomic Nuclear. *Unnes Science Education Journal*, *6*(2), 1583–1590. https://doi.org/10.15294/usej.v6i2.15856