ISEJ Indonesian Science Education Journal

ISSN Online: 2716-3350
Vol. 4 No. 1 January 2023 Pages 10-18
https://siducat.org/index.php/isej/
Published by Yayasan Darussalam Bengkulu

Development and Validation of a Two-Tier Multiple Choice Instrument to Assess Students' Science Process Skills in Acid-Base Chemistry

Meri Andria[1], Hadiwinarto[2]
Universitas Bengkulu[1,2]
E-mail: mery_swaraunibfm@yahoo.com

Abstract
This study aims to develop and validate a two-tier multiple choice (TTMCQ) instrument to measure senior high school students' science process skills (SPS) in acid-base chemistry. Following Borg and Gall's modified research and development design (nine stages), the instrument was developed, validated, and tested in two public high schools in Bengkulu, Indonesia. Validation involved content experts, assessment specialists, and experienced chemistry teachers. Field trials were conducted with 119 students, and data were analyzed for validity, reliability, item difficulty, and discrimination indices. Results show that the TTMCQ instrument demonstrated strong validity, high reliability, appropriate difficulty distribution (53% easy, 42% medium, 5% difficult), and sufficient discriminating power. Teachers reported positive perceptions, noting that the instrument effectively challenges students' reasoning and enhances their SPS. The study highlights the potential of TTMCQ instruments as practical tools for formative and summative assessment in science education.
Keywords: **Assessment Instrument; Science Process Skills; Two-Tier Multiple Choice Question.**

INTRODUCTION

Teachers play a fundamental role in guiding, facilitating, and evaluating students' learning processes in the classroom. Beyond the responsibility of delivering subject matter, teachers are expected to design and implement effective assessment strategies that provide meaningful insights into students' progress and achievements (Almoslamani, 2022; Veldhuis & van den Heuvel-Panhuizen, 2020; Yan et al., 2021). Assessment is not merely the act of assigning grades but a systematic process of collecting, analyzing, and interpreting data about students' learning outcomes to support pedagogical decision-making and instructional improvement (Schellekens et al., 2021; Schildkamp et al., 2020; Thornhill-Miller et al., 2023). Effective assessment allows teachers to determine the extent of learning objectives achieved, identify student misconceptions, monitor the effectiveness of instructional strategies, and classify students' mastery levels in a fair and evidence-based manner (Hill & Chin, 2018; Simonson et al., 2022; Sun et al., 2023).

Science education, in particular, requires a comprehensive assessment approach that addresses both conceptual understanding and the processes by which scientific knowledge is constructed. Science is not only a body of knowledge but also a process of inquiry involving exploration, experimentation, and reasoning (Nur'ariyani et al., 2023; Özer & Sarıbaş, 2023; Zhou et al., 2016). To engage in authentic scientific inquiry, students must develop science process skills (SPS), which encompass the cognitive, manipulative, and social competencies necessary to construct knowledge, solve problems, and communicate findings (Gizaw & Sota, 2023; Ilma et al., 2020; Nugroho et al., 2023). SPS are typically categorized into basic skills such as observing, classifying, measuring, inferring, and predicting and integrated skills, including hypothesizing, controlling variables, designing experiments, interpreting data, and drawing conclusions (Astalini et al., 2023; Ekici & Erdem, 2020; Rusmini et al., 2021). The mastery of SPS is widely acknowledged as a cornerstone of science education, as it fosters critical thinking, creativity, and scientific literacy among students (Primasari et al., 2020; Rahmatika et al., 2022; Sidek et al., 2020).

Chemistry, as a branch of science, poses unique challenges in teaching and learning due to its abstract concepts, symbolic representations, and experimental orientation. Learning chemistry effectively requires not only memorization of facts but also the ability to connect macroscopic phenomena, submicroscopic structures, and symbolic models a framework known as the chemistry triplet (Permatasari et al., 2022; Rahmawati et al., 2022; Uyulgan & Güven, 2022). To address these challenges, chemistry education should emphasize active engagement in scientific practices through the development of SPS (Adauyah & Aznam, 2024; Ahmad & Iksan, 2021; Purtadi et al., 2023). Research has shown that students often struggle with acid-base concepts due to their complexity and abstractness, which underscores the importance of developing valid assessment instruments that reveal not only what students know but also how they reason about their knowledge (Damanhuri et al., 2016; Grogans et al., 2023; Salame et al., 2022).

Traditional assessment methods, such as multiple-choice questions (MCQs), are widely used because they are practical and efficient. However, conventional MCQs often fail to capture the depth of students' conceptual understanding and their reasoning processes, allowing for guessing and superficial learning (Holzinger et al., 2020; Jovanovska, 2018; Saher et al., 2022). To overcome these limitations, researchers have proposed the use of two-tier multiple choice questions (TTMCQs), which consist of a content-based first tier and a reasoning-based second tier (Lukum et al., 2023; Rintayati et al., 2021; Sıbıç et al., 2022). TTMCQs have been shown to be more effective than traditional MCQs in diagnosing misconceptions, promoting higher-order thinking, and assessing SPS in science education (El Nagdi & Roehrig, 2022; Kasule, 2013; Maison et al., 2022). Recent studies further demonstrate that TTMCQs can enhance students' engagement with content and provide teachers with richer diagnostic information to support instructional design (Alioon & Delialioğlu, 2019; Boekaerts, 2016; Bond & Bedenlier, 2019).

Despite these advances, the application of TTMCQs in chemistry education, particularly in assessing SPS in specific topics such as acid-base concepts, remains underexplored. Previous studies have focused on the use of TTMCQs for identifying misconceptions in thermochemistry, equilibrium, or genetics (Jenkins & Shoopman, 2019), but little attention has been given to their role in systematically measuring SPS. Moreover, few studies have validated TTMCQ instruments in the context of Indonesian secondary education, where challenges in meeting national assessment standards persist (Haekal, 2022; Misbah et al., 2020; Soeharto & Csapó, 2022). This gap highlights the need for a rigorously developed and validated assessment tool that not only addresses students' conceptual understanding but also captures their SPS across acid-base learning materials.

Therefore, the present study aims to develop and validate a two-tier multiple choice assessment instrument specifically designed to measure senior high school students' SPS on acid base chemistry. The study seeks to establish the psychometric properties of the instrument including validity, reliability, difficulty level, and discrimination index while also exploring teachers' responses to its feasibility and practicality in classroom application. By addressing this gap, the study contributes to the advancement of assessment practices in science education and provides an alternative tool for promoting inquiry-based learning and improving the quality of chemistry instruction at the secondary school level.

METHODS

This study employed a research and development (R&D) design that was adapted from the Borg and Gall model, which is widely recognized in educational research for guiding systematic product development. To maintain feasibility within the scope of the study, nine stages of the model were implemented, including preliminary research and information gathering, planning, initial product design, preliminary field testing, revision of the initial product, field testing, revision based on field results, operational testing, and final product revision. The tenth stage, dissemination and implementation, was excluded due to its extensive scope and resource requirements. This stepwise process ensured that the developed assessment instrument underwent iterative refinement informed by both expert evaluation and empirical testing.

The research was conducted in two Indonesian public senior high schools, SMA N 1 Bengkulu and SMA N 2 Bengkulu, during the 2014/2015 academic year. Participants included 119 students and four chemistry teachers selected through purposive sampling to represent diverse academic abilities and teaching experiences. Students were distributed across different trial phases, including small-scale and large-scale testing, to minimize learning effects from repeated exposure to similar test items. The involvement of teachers as participants was intended to capture professional perspectives on the feasibility and practicality of the developed instrument. Ethical considerations were addressed through informed consent, voluntary participation, and anonymity in data reporting.

The product developed in this study was a two-tier multiple choice question (TTMCQ) instrument designed to measure students' science process skills (SPS) in acid-base chemistry. The first tier of each item consisted of a statement requiring a factual or conceptual response, while the second tier elicited the reasoning behind the chosen answer. The inclusion of the reasoning tier was aimed at reducing guessing and allowing for a more accurate assessment of students' SPS. The instrument development process began with the construction of a blueprint aligned with the national curriculum, learning objectives, and indicators of SPS, including observing, classifying, predicting, hypothesizing, planning experiments, interpreting data, and communicating scientific ideas. The initial draft consisted of item grids, test items, answer keys, and scoring rubrics.

Expert validation was conducted to ensure both content and construct validity. Material experts in acid-base chemistry reviewed the items for accuracy, clarity, and alignment with curricular standards, while assessment specialists evaluated the construct validity of the instrument to confirm its ability to measure SPS. In addition, senior chemistry teachers were consulted to provide feedback on the practicality and classroom relevance of the items. Revisions were made based on their suggestions, such as improving item clarity, adjusting distractors, and refining the wording of instructions.

The revised instrument underwent pilot testing through one-on-one trials with six students and small-group trials with 14 students and two teachers. These trials assessed readability, clarity of instructions, and students' perceptions of item difficulty. Field testing followed with a larger sample of 53 students and two teachers, allowing for item analysis to determine validity, reliability, item difficulty, and discrimination indices. Reliability was measured using Cronbach's alpha to ensure internal consistency, while item validity was established through expert judgment and statistical correlation analysis. The results guided further revisions before conducting operational testing with 66 students and two teachers. The operational trial assessed the instrument's practicality and feasibility for large-scale classroom implementation without direct researcher assistance.

Data collection methods included questionnaires, observations, and test results. Teacher and student responses to the instrument were obtained using structured questionnaires, while classroom observations provided additional qualitative insights into practicality. Quantitative data were analyzed using descriptive and inferential statistics. Item difficulty was categorized into easy, moderate, and difficult levels, while discrimination indices classified the effectiveness of items in distinguishing between high- and low-performing students. A two-tailed t-test was employed to compare the performance of students assessed using the TTMCQ instrument with those assessed using conventional multiple choice questions, thereby examining the effectiveness of the developed product.

Through this systematic methodology, the study ensured the validity, reliability, and practicality of the TTMCQ instrument as a tool for assessing science process skills in secondary school chemistry. The use of iterative development, expert validation, and empirical testing provided a robust foundation for establishing the instrument's feasibility in authentic classroom contexts.

RESULTS AND DISCUSSION

Characteristics of the Developed TTMCQ Instrument

The two-tier multiple choice question (TTMCQ) instrument developed in this study was designed to measure students' science process skills (SPS) in acid-base chemistry. The instrument incorporated indicators such as observing, classifying, predicting, hypothesizing, planning experiments, interpreting data, and communicating. Each item contained a first-tier question assessing factual or conceptual knowledge and a second-tier question eliciting the reasoning behind students' responses. Expert validation confirmed that the items demonstrated strong content validity, with acid-base material experts ensuring conceptual accuracy, and assessment experts affirming construct alignment with SPS. Senior chemistry teachers further evaluated the practicality of the instrument, noting that the two-tier format encouraged deeper thinking compared to conventional multiple choice questions.

Validity and Reliability Analysis

The statistical analysis of the field trials revealed that the TTMCQ instrument achieved acceptable psychometric properties. Item validity was confirmed through expert judgment and correlation tests, while internal consistency reliability, measured by Cronbach's alpha, was found to be high. The distribution of item difficulty showed that 53% of items were categorized as easy, 42% as moderate, and 5% as difficult, representing a balanced spread across levels. The discrimination index indicated that most items had sufficient power to differentiate between high- and low-achieving students. These results suggest that the TTMCQ instrument meets established standards for educational assessment tools.

Teacher and Student Responses

Teachers' responses to the instrument were mixed but generally positive. While they acknowledged that two-tier questions were more complex to construct and administer than conventional multiple choice questions, they agreed that the TTMCQ format effectively challenged students' reasoning and minimized the likelihood of guessing. Students reported that the instrument required more time and cognitive effort but simultaneously stimulated critical analysis of concepts and reasoning. Both groups emphasized the novelty of the instrument in classroom practice, as it had not previously been implemented in their schools.

Comparison with Conventional MCQs

The effectiveness of the TTMCQ instrument was examined through a comparison with conventional multiple choice questions (MCQs) using a two-tailed t-test. Results indicated no significant difference in mean scores between the two groups, suggesting that students were equally capable of working with either format. However, qualitative responses revealed that the TTMCQ provided richer diagnostic information regarding reasoning and conceptual understanding, a dimension often absent in standard MCQs.

Discussion

The findings of this study align with prior research emphasizing the benefits of TTMCQs in diagnosing misconceptions and promoting higher-order thinking. Yeo et al. (2022), for instance, demonstrated that TTMCQs are superior to traditional MCQs in identifying alternative conceptions in science learning. Darling-Hammond et al. (2020) found that the inclusion of reasoning tiers allows teachers to better assess students' conceptual depth, which resonates with the responses of teachers in the present study. The positive reception by teachers and students also supports Williams (2024) findings that TTMCQs enhance formative assessment practices by encouraging more reflective engagement with content.

In chemistry education specifically, studies by Karpudewan (2016), Dejene and Belachew (2023) highlight the role of TTMCQs in improving students' conceptual understanding in topics such as equilibrium and reaction rates. The current study extends these findings by applying TTMCQs to acid-base concepts and explicitly linking them to SPS indicators, thereby addressing a gap in previous research. This novelty lies in positioning TTMCQs not only as diagnostic tools but also as structured instruments for measuring SPS systematically, a contribution that has not been extensively documented in earlier studies.

Furthermore, the balanced difficulty distribution observed in this study echoes the recommendations of Schildkamp (2020), who emphasized that well-constructed assessments should provide a range of difficulty levels to capture diverse student abilities. The high reliability score of the instrument also corroborates findings from Gizaw and Sota (2023), who argued that reliable assessments are essential for accurately measuring SPS and informing instructional practices.

The implications of this study are twofold. First, for educators, the TTMCQ instrument provides a practical alternative to conventional MCQs by offering insights into both knowledge and reasoning, thereby supporting more meaningful formative and summative assessments. Second, for curriculum developers and policymakers, the findings suggest that integrating TTMCQs into national assessments could enhance the alignment between instructional practices and the development of SPS, which are critical competencies in 21st-century science education.

Despite these strengths, the study has several limitations. The research was conducted in only two schools within one regional context, which may limit the generalizability of findings to broader educational settings. The instrument was also applied only to acid-base chemistry, leaving open the question of whether similar results would be achieved in other scientific domains. Additionally, while teacher and student perceptions were captured, long-term impacts on learning outcomes were not examined. Future studies should therefore extend the use of TTMCQs across different scientific topics, include larger and more diverse populations, and investigate longitudinal effects on SPS development.

CONCLUSION

This study successfully developed and validated a two-tier multiple choice question (TTMCQ) instrument designed to measure senior high school students' science process skills (SPS) in acid-base chemistry, demonstrating strong content and construct validity, high reliability, balanced item difficulty, and sufficient discriminating power. Teacher and student responses indicated that, although more cognitively demanding than conventional multiple choice questions, the TTMCQ format effectively stimulated reasoning, minimized guessing, and provided richer diagnostic insights into students' conceptual understanding and SPS development. The novelty of this study lies in systematically integrating TTMCQs with SPS indicators in the context of acid-base chemistry, thereby extending prior research that predominantly focused on diagnosing misconceptions in other scientific domains. The findings imply that TTMCQs can serve as a practical and reliable assessment tool to support inquiry-based learning, formative evaluation, and instructional improvement in secondary science education. While the study was limited to two schools and one topic area, its results offer a promising foundation for broader applications of TTMCQs across diverse scientific subjects and educational contexts, highlighting their potential contribution to strengthening assessment practices and advancing science education quality.

REFERENCE

Adauyah, R., & Aznam, N. (2024). Guided Inquiry Learning Model in Chemistry Education: A Systematic Review. *Jurnal Penelitian Pendidikan IPA, 10*(3), 77-87. https://doi.org/10.29303/jppipa.v10i3.6373

Ahmad, N. F., & Iksan, Z. H. (2021). Edmodo-Based Science Module Development on Students' Mastery of Science Process Skills: Need Analysis. *Creative Education, 12*(11), 2609-2623. https://doi.org/10.4236/ce.2021.1211195

Alioon, Y., & Delialioğlu, Ö. (2019). The effect of authentic m-learning activities on student engagement and motivation. *British Journal of Educational Technology, 50*(2), 655-668. https://doi.org/10.1111/bjet.12559

Almoslamani, Y. (2022). The impact of learning strategies on the academic achievement of university students in Saudi Arabia. *Learning and Teaching in Higher Education: Gulf Perspectives, 18*(1), 4-18. https://doi.org/10.1108/LTHE-08-2020-0025

Astalini, A., Darmaji, D., Kurniawan, D. A., Wirayuda, R. P., Putri, W. A., Rini, E. F. S., Ginting, A. A. B., & Ratnawati, T. (2023). Impact of Science Process Skills on Thinking Skills in Rural and Urban Schools. *International Journal of Instruction, 16*(2), 803-822. https://doi.org/10.29333/iji.2023.16242a

Boekaerts, M. (2016). Engagement as an inherent aspect of the learning process. *Learning and Instruction, 43*, 76-83. https://doi.org/10.1016/j.learninstruc.2016.02.001

Bond, M., & Bedenlier, S. (2019). Facilitating Student Engagement Through Educational Technology: Towards a Conceptual Framework. *Journal of Interactive Media in Education, 2019*(1). https://doi.org/10.5334/jime.528

Damanhuri, M. I. M., Treagust, D. F., Won, M., & Chandrasegaran, A. L. (2016). High School Students' Understanding of Acid-Base Concepts: An Ongoing Challenge for Teachers. *The International Journal of Environmental and Science Education.* https://doi.org/10.12973/ijese.2015.284a

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97-140. https://doi.org/10.1080/10888691.2018.1537791

Dejene, K., & Belachew, W. (2023). Topic-Specific Pedagogical Content Knowledge-Based Instruction and Level of Conceptual Understanding of Chemical Kinetics and Equilibrium concepts on Grade 11 Students. *Science Education International, 34*(2), 96-108. https://doi.org/10.33828/sei.v34.i2.3

Ekici, M., & Erdem, M. (2020). Developing Science Process Skills through Mobile Scientific Inquiry. *Thinking Skills and Creativity, 36*, 100658. https://doi.org/10.1016/j.tsc.2020.100658

El Nagdi, M. A., & Roehrig, G. H. (2022). Reality vs. Expectations of Assessment in STEM Education: An Exploratory Case Study of STEM Schools in Egypt. *Education Sciences, 12*(11), 762. https://doi.org/10.3390/educsci12110762

Gizaw, G., & Sota, S. S. (2023). Improving Science Process Skills of Students: A Review of Literature. *Science Education International, 34*(3), 216-224. https://doi.org/10.33828/sei.v34.i3.5

Grogans, S. E., Bliss-Moreau, E., Buss, K. A., Clark, L. A., Fox, A. S., Keltner, D., Cowen, A. S., Kim, J. J., Kragel, P. A., MacLeod, C., Mobbs, D., Naragon-Gainey, K., Fullana, M. A., & Shackman, A. J. (2023). The nature and neurobiology of fear and anxiety: State of the science and opportunities for accelerating discovery. *Neuroscience & Biobehavioral Reviews, 151*, 105237. https://doi.org/10.1016/j.neubiorev.2023.105237

Haekal, M. (2022). Tantangan Distribusi Guru di Daerah Terpencil Indonesia: Antara Manajemen, Isu Personal, dan Faktor Geografi. *TA'DIB: Jurnal Pemikiran Pendidikan, 12*(1), 18-23. https://doi.org/10.54604/tdb.v12i1.100

Hill, H. C., & Chin, M. (2018). Connections Between Teachers' Knowledge of Students, Instruction, and Achievement Outcomes. *American Educational Research Journal, 55*(5), 1076-1112. https://doi.org/10.3102/0002831218769614

Holzinger, A., Lettner, S., Steiner-Hofbauer, V., & Capan Melser, M. (2020). How to assess? Perceptions and preferences of undergraduate medical students concerning traditional assessment methods. *BMC Medical Education, 20*(1), 312. https://doi.org/10.1186/s12909-020-02239-6

Ilma, S., Al-Muhdhar, M. H. I., Rohman, F., & Saptasari, M. (2020). The correlation between science process skills and biology cognitive learning outcome of senior high school students. *JPBI (Jurnal Pendidikan Biologi Indonesia, 6*(1), 55-64. https://doi.org/10.22219/jpbi.v6i1.10794

Jenkins, J. L., & Shoopman, B. T. (2019). Identifying misconceptions that limit student understanding of molecular orbital diagrams. *Science Education International, 30*(3), 152-157. https://doi.org/10.33828/sei.v30.i3.1

Jovanovska, J. (2018). Designing Effective Multiple-Choice Questions for Assessing Learning Outcomes. *Infotheca, 18*(1), 25-42. https://doi.org/10.18485/infotheca.2018.18.1.2

Karpudewan, M., Treagust, D. F., Mocerino, M., Won, M., & Chandrasegaran, A. L. (2016). Investigating High School Students' Understanding of Chemical Equilibrium Concepts. *The International Journal of Environmental and Science Education.* https://doi.org/10.12973/ijese.2015.280a

Kasule, O. H. (2013). Overview of medical student assessment: Why, what, who, and how. *Journal of Taibah University Medical Sciences, 8*(2), 72-79. https://doi.org/10.1016/j.jtumed.2013.07.001

Lukum, A., Karongkong, N., Pikoli, M., Tangio, J. S., Mohamad, E., & Rewini Kunusa, W. (2023). Analysis of Student's Conceptual Understanding Using Two-Tier Multiple Choice Diagnostic Test on Acid-Base Topic. *E3S Web of Conferences, 400*, 04003. https://doi.org/10.1051/e3sconf/202340004003

Maison, M., Hidayat, M., Kurniawan, D. A., Yolviansyah, F., Sandra, R. O., & Iqbal, M. (2022). How Critical Thinking Skills Influence Misconception in Electric Field. *International Journal of Educational Methodology, 8*(2), 377-390. https://doi.org/10.12973/ijem.8.2.377

Misbah, Z., Gulikers, J., Dharma, S., & Mulder, M. (2020). Evaluating competence-based vocational education in Indonesia. *Journal of Vocational Education & Training, 72*(4), 488-515. https://doi.org/10.1080/13636820.2019.1635634

Nugroho, I., Aufa, M., Chasanah, N. N., & Wardani, D. A. (2023). Analysis of Higher Order Thinking Skills and Science Process Skills in 21st Century Education (pp. 729-735). https://doi.org/10.2991/978-94-6463-284-2_78

Nur'ariyani, S., Jumyati, J., Yuliyanti, Y., Nulhakim, L., & Leksono, S. M. (2023). Scientific Approach to Learning Science in Elementary Schools. *Jurnal Penelitian Pendidikan IPA, 9*(8), 6659-6666. https://doi.org/10.29303/jppipa.v9i8.3680

Özer, F., & Sarıbaş, D. (2023). Exploring Pre-service Science Teachers' Understanding of Scientific Inquiry and Scientific Practices Through a Laboratory Course. *Science & Education, 32*(3), 787-820. https://doi.org/10.1007/s11191-022-00325-3

Permatasari, M. B., Rahayu, S., & Dasna, I. W. (2022). Chemistry Learning Using Multiple Representations: A Systematic Literature Review. *Journal of Science Learning, 5*(2), 334-341. https://doi.org/10.17509/jsl.v5i2.42656

Primasari, R., Miarsyah, M., & Rusdi, R. (2020). Science literacy, critical thinking skill, and motivation: A correlational study. *JPBI (Jurnal Pendidikan Biologi Indonesia, 6*(2), 273-282. https://doi.org/10.22219/jpbi.v6i2.11124

Purtadi, S., Suyanta, S., & Rohaeti, E. (2023). Science Process Skills as Learning Outcomes in Inquiry-Based Learning: A Literature Review. *Jurnal Penelitian Pendidikan IPA, 9*(12), 1403-1414. https://doi.org/10.29303/jppipa.v9i12.4774

Rahmatika, R., Amin, M., Al-Muhdhar, M. H. I., & Suwono, H. (2022). Scientific literacy refinement at Islamic junior high schools using socio-science spirituality learning model. *JPBI (Jurnal Pendidikan Biologi Indonesia, 8*(1), 40-50. https://doi.org/10.22219/jpbi.v8i1.18989

Rahmawati, Y., Zulhipri, Z., Hartanto, O., Falani, I., & Iriyadi, D. (2022). Students' conceptual understanding in chemistry learning using PhET interactive simulations. *Journal of Technology and Science Education, 12*(2), 303. https://doi.org/10.3926/jotse.1597

Rintayati, P., Lukitasari, H., & Syawaludin, A. (2021). Development of Two-Tier Multiple Choice Test to Assess Indonesian Elementary Students' Higher-Order Thinking Skills. *International Journal of Instruction, 14*(1), 555-566. https://doi.org/10.29333/iji.2021.14133a

Rusmini, R., Suyono, S., & Agustini, R. (2021). Analysis of science process skills of chemical education students through self project based learning (SjBL) in the pandemic COVID 19 era. *Journal of Technology and Science Education, 11*(2), 371. https://doi.org/10.3926/jotse.1288

Saher, A.-S., Ali, A. M. J., Amani, D., & Najwan, F. (2022). Traditional versus authentic assessments in higher education. *Pegem Journal of Education and Instruction, 12*(1). https://doi.org/10.47750/pegegog.12.01.29

Salame, I. I., Ramirez, L., Nikolic, D., & Krauss, D. (2022). Investigating Students' Difficulties and Approaches to Solving Buffer Related Problems. *International Journal of Instruction, 15*(1), 911-926. https://doi.org/10.29333/iji.2022.15152a

Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation, 71*, 101094. https://doi.org/10.1016/j.stueduc.2021.101094

Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research, 103*, 101602. https://doi.org/10.1016/j.ijer.2020.101602

Sidek, R., Halim, L., Buang, N. A., & Mohamad Arsad, N. (2020). Fostering Scientific Creativity in Teaching and Learning Science in Schools: A Systematic Review. *Jurnal Penelitian Dan Pembelajaran IPA, 6*(1), 13. https://doi.org/10.30870/jppi.v6i1.7149

Simonson, S. R., Earl, B., & Frary, M. (2022). Establishing a Framework for Assessing Teaching Effectiveness. *College Teaching, 70*(2), 164-180. https://doi.org/10.1080/87567555.2021.1909528

Sıbıç, O., AKÇAY, B., & ARIK, M. (2022). Review of Two-tier Tests in the Studies: Creating a New Pathway for Development of Two-tier Tests. *International Journal of Contemporary Educational Research, 7*(2), 81-98. https://doi.org/10.33200/ijcer.747981

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian student misconceptions in science concepts. *Heliyon, 8*(9), e10720. https://doi.org/10.1016/j.heliyon.2022.e10720

Sun, L., Kangas, M., Ruokamo, H., & Siklander, S. (2023). A systematic literature review of teacher scaffolding in game-based learning in primary education. *Educational Research Review, 40*, 100546. https://doi.org/10.1016/j.edurev.2023.100546

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence, 11*(3). https://doi.org/10.3390/jintelligence11030054

Uyulgan, M. A., & Güven, N. A. (2022). Analysis of Prospective Primary School Teachers' Knowledge Regarding Chemical Representations on Crystallization Experiment. *Journal of Science Learning, 5*(1), 176-192. https://doi.org/10.17509/jsl.v5i1.34772

Veldhuis, M., & van den Heuvel-Panhuizen, M. (2020). Supporting primary school teachers' classroom assessment in mathematics education: effects on student achievement. *Mathematics Education Research Journal, 32*(3), 449-471. https://doi.org/10.1007/s13394-019-00270-5

Williams, A. (2024). Delivering Effective Student Feedback in Higher Education: An Evaluation of the Challenges and Best Practice. *International Journal of Research in Education and Science, 10*(2), 473-501. https://doi.org/10.46328/ijres.3404

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice, 28*(3), 228-260. https://doi.org/10.1080/0969594X.2021.1884042

Yeo, J.-H., Yang, H.-H., & Cho, I.-H. (2022). Using A Three-Tier Multiple-Choice Diagnostic Instrument Toward Alternative Conceptions Among Lower-Secondary School Students In Taiwan: Taking Ecosystems Unit as An Example. *Journal of Baltic Science Education, 21*(1), 69-83. https://doi.org/10.33225/jbse/22.21.69

Zhou, S., Han, J., Koenig, K., Raplinger, A., Pi, Y., Li, D., Xiao, H., Fu, Z., & Bao, L. (2016). Assessment of scientific reasoning: The effects of task context, data, and design on student reasoning in control of variables. *Thinking Skills and Creativity, 19*, 175-187. https://doi.org/10.1016/j.tsc.2015.11.004