EduChemia Vol.10, No.1, 2025, p. 115-128
Jurnal Kimia dan Pendidikan e-ISSN 2502-4787

CONTENT VALIDITY OF ASSESSMENT INSTRUMENT AND STORYBOARD TWO-TIER VIRTUAL REALITY (TTVR) IN ACID BASE MATERIAL

Yuli Astuti[1], Sri Yamtinah[1*], Sulistyo Saputro[1], Ari Syahidul Shidiq[1], Sudarmin[2], Sri Rahayu[3]

[1]*Department of Chemistry Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Surakarta, Indonesia.*
[2]*Department of Chemistry Education, Faculty of Mathematics and Natural Science, Universitas Negeri Semarang, Semarang, Indonesia.*
[3]*Department of Chemistry Education, Faculty of Mathematics and Natural Science, Universitas Negeri Malang, Malang, Indonesia.*

E-mail: *jengtina@staff.uns.ac.id

Received: 28 Mei 2025. Accepted: 9 Juli 2025. Published: 15 Juli 2025

DOI: 10.30870/educhemia.v10i1. 33166

Abstract: The aim of this research is to analyze the content validity of the TTVR assessment instrument for acid–base materials based on the Aiken validity index. This ensures that the instrument can effectively measure what it intends to assess. Content validity data were obtained from seven experts, including two chemistry education lecturers from UNS and five high school teachers in Surakarta. The content validity test used a 1–4 rating scale, with categories defined as follows: relevant with a score of 4; less relevant with a score of 3; irrelevant with a score of 2; and quite relevant with a score of 1. This scale was applied to 15 instrument questions and 16 storyboard indicators related to media and material aspects. The development results of the TTVR assessment instrument for acid–base content achieved an Aiken index of over 0.76, indicating that the instrument is valid. This suggests that the TTVR assessment instrument and storyboard are suitable, with improvements made based on suggestions and feedback from experts.

Keywords: Aiken; Storyboard; Two-tier; Validity; Virtual reality

Abstrak: Tujuan penelitian ini adalah menganalisis validitas isi instrumen asesmen TTVR untuk materi asam-basa berdasarkan indeks validitas Aiken. Hal ini memastikan bahwa instrumen tersebut dapat mengukur secara efektif apa yang ingin dinilai. Data validitas isi diperoleh dari tujuh pakar, termasuk dua dosen pendidikan kimia dari UNS dan lima guru SMA di Surakarta. Uji validitas isi menggunakan skala penilaian 1-4, dengan kategori yang didefinisikan sebagai berikut: relevan dengan skor 4; kurang relevan dengan skor 3; tidak relevan dengan skor 2; dan cukup relevan dengan skor 1. Skala ini diterapkan pada 15 pertanyaan instrumen dan 16 indikator storyboard yang terkait dengan aspek media dan materi. Hasil pengembangan instrumen asesmen TTVR untuk materi asam-basa mencapai indeks Aiken lebih dari 0,76, yang menunjukkan bahwa instrumen tersebut valid. Hal ini menunjukkan bahwa instrumen penilaian TTVR dan storyboard dianggap sesuai, dengan perbaikan yang dilakukan berdasarkan saran dan masukan dari para ahli.

Kata kunci: Aiken, Storyboard, Two-tier, Validitas, Virtual reality

INTRODUCTION

Chemistry is a subject that focuses on understanding the structure, properties, laws, principles, theories and concepts that explain changes in matters. The characteristics of chemistry tend to be abstract, interconnected, related to calculations, and not directly observable (Santi & Rahayu, 2022). As a result, students often perceive chemistry as a difficult subject to understand. However, chemistry can be more easily comprehended by representing it at three levels: macroscopic, submicroscopic, and symbolic (Chandrasegaran *et al*., 2007)

Macroscopic representation is the representation that is directly observed through the senses. Submicroscopic representation is abstract and depicts the structure, molecules, atoms, or ions of matter. Symbolic representations include formulas, chemical reaction equations, mathematical calculations, and symbols of matter. Learning that involves the interconnection of these three levels of representation is an important aspect that teachers must pay attention to (Santi & Rahayu, 2022).

However, in reality, the learning conditions in the field often differ. Observations conducted at SMAN 3 Boyolali indicate that chemistry instruction, particularly on acids and bases, relies heavily on lectures and practical methods. These methods focus primarily on the macroscopic and symbolic levels, whereas the submicroscopic level is neglected (Santi & Rahayu, 2022). This imbalance in the learning process can lead to students being less effective in achieving a deeper understanding of concepts (Setyoko *et al*., 2018)

An imbalance in understanding can lead to difficulties for students in grasping the material, which in turn decreases their academic performance and motivation to learn. Consequently, the potential for misconceptions in understanding the material increases. Misconceptions arise from an inaccurate understanding of information, hindering students' ability to answer questions correctly (Gradini, 2016; Saputri *et al*., 2016). For example, for acids and bases, students mistakenly classify H_2S, a weak acid, as a strong acid (Utami *et al*., 2020)

Early diagnosis of misconceptions is crucial for assessing students' understanding through diagnostic tests, which can effectively identify areas of misunderstanding (Mutmainna *et al*., 2018). Commonly used diagnostic tests include multiple-choice and essay formats (Damayanti *et al*., 2018). This aligns with the results of a needs analysis conducted through interviews with teachers at SMAN 3, Boyolali.

Although multiple-choice tests can assess knowledge, they are not effective in measuring problem-solving skills (Wartoni & Benyamin, 2020). In contrast, essay tests allow for more personal responses but can be difficult to evaluate objectively (Supardi, 2015). Therefore, alternative diagnostic tests that can evaluate students' understanding while also fostering higher-order thinking skills are needed (Hidayatullah *et al*., 2022).

In this study, a two-tier multiple-choice (TTMC) assessment instrument was developed. The TTMC consists of questions with two levels: the first level presents the main answer, whereas the second level provides the reasoning (Chandrasegaran *et al*., 2007). With TTMC, students can be distinguished between those who answer correctly with incorrect reasoning and those who answer correctly with correct reasoning. Research related to diagnosing misconceptions using TTMC has been conducted by various researchers, including studies on misconceptions in chemical bonding (Gultom *et al*., 2023); concepts, theories, and calculations of acids and bases (Fajri *et al*., 2020); and concepts of electrolytic and nonelectrolytic solutions (Siswaningsih *et al*., 2015). This test is widely chosen because it is effective and efficient in identifying misconceptions among students (Utami *et al*., 2020)

TTMC can measure only students' cognitive understanding of macroscopic and symbolic aspects, necessitating an alternative approach that focuses on the submicroscopic level, such as virtual reality (VR). VR is a technology that allows users to interact within a three-dimensional (3D) virtual environment (Jamil, 2018). The use of VR as a supporting tool for research instruments has not been extensively explored.

In this study, VR is integrated with culture through ethnochemistry, specifically the ritual of cleansing keris in the context of acids and bases called two-tier multiple choice with VR (TTVR). With VR, the concepts of acids and bases can hopefully be visualized according to the three representative levels of chemistry learning. This combination creates a deep and meaningful learning assessment experience, thereby enhancing student engagement in the learning process.

This research focuses on the validation of the TTMC instrument and VR storyboard through expert evaluation and the Aiken index (Aiken, 1985). Thus, this study aims to provide a valid assessment instrument that bridges ethnochemical and innovative approaches in chemistry education.

This study does not include the implementation and evaluation stages; instead, it focuses on developing the assessment instrument. This approach ensures the instrument's validity and effectiveness before broader application, aiming to address the weaknesses of the instrument. This study provides a solid foundation for future research on its application in learning processes.

METHOD

The method used in this study is research and development (R&D) (Sugiyono, 2016). This study developed an integrated TTVR assessment instrument for acid–base materials via the ADDIE development model created by Dick & Carey (1996). The ADDIE model stands for Analyze, Design, Development, Implementation, and Evaluation (Susanto & Ayuni, 2017).

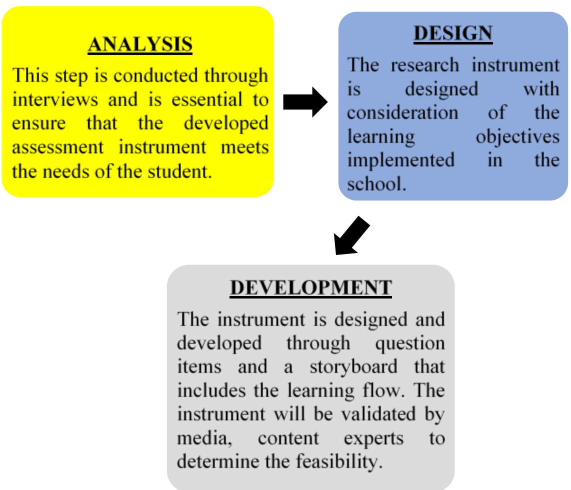

Figure 1. ADDIE Model

Sample and Data Collection

To assess the extent to which a measurement instrument can perform its measurement function, validity testing is conducted. An instrument that yields high validity means that it provides measurement results that align with the objectives of the measurement. This study measures validity through the content validity of the assessment instrument related to the developed material and media (Sudaryono *et al*., 2019). Content validity refers to validity evaluated from the perspective of the test's content as a tool for assessing learning outcomes.

Validity testing was conducted through focus group discussions (FGDs) with lecturers from Sebelas Maret University (UNS) and high school chemistry teachers from Surakarta. The validation focused on the items of the assessment instrument and the storyboard. Validation was carried out by assessing the alignment of the question indicators with the TTVR-ethnochemical indicators, the appropriateness of the concepts, the content of the question items and the VR storyboard, the suitability of the answer keys, and the choice of words in each question item and VR storyboard.

Data Analysis

The validity analysis of the content of the TTVR-Ethnochemistry assessment instrument on acid–base materials in this study uses the Aiken formula with the following formula (Aiken, 1980):

$$V = \sum S/[n(C-1)]$$

$$S = R - Lo$$

Information:

V = Aiken index

S = score given by the assessor minus the lowest score in the category

R = score given by the assessor

Lo = lowest assessment score (1)

C = highest assessment score (4)

n = number of validators (assessors)

The assessment was conducted using a questionnaire with a Likert scale ranging from 1–4. This scale avoids neutral or uncertain ratings, so the questionnaire was designed with a composition of both positive and negative items (Sugiyono, 2016). The TTVR-Ethnochemistry assessment instrument and storyboard on acid–base materials are considered good if their validity is at least acceptable, with a V value. Therefore, a V value on the 1–4 Likert scale is deemed good if it exceeds 0.76 (Aiken, 1985). This value indicates that the TTVR-Ethnochemistry assessment instrument is usable and relevant for measuring the variables under investigation. The Likert scale is presented in Table 1. The results of the data obtained are then used to improve product development so that the results of the educational products that have been tested are valid and suitable for use in the learning process.

Table 1. Assessment criteria and Likert scale score acquisition

Student Answer Type	Score
Relevant	4
Quite Relevant	3
Less Relevant	2
Irrelevan	1

RESULTS AND DISCUSSION

Analysis

Needs analysis is the initial stage conducted through interviews with teachers and students regarding the evaluation of learning outcomes. Based on interviews conducted with three high school chemistry teachers in Surakarta, common issues identified among students include difficulties in understanding abstract chemistry concepts. The teaching methods in the classroom are predominantly dominated by lectures, discussions, and practical work (Sokpheng & Meng, 2022). The limitations of visual representation in learning often cause students to struggle with deepening their understanding of the material. This is related to the difficulty of visualizing chemical concepts at the macroscopic, submicroscopic, and symbolic levels (Azizah *et al*., 2022).

Based on the interview results, the common issue that arises is related to the assessment instruments used to measure the level of conceptual understanding. In the classroom, nontest assessment instruments include individual assignments, practical reports, and tests in the form of multiple-choice questions or essay questions. The assessment instruments that have been used are considered insufficiently effective in exploring students' deep understanding of material concepts (Desiriah & Setyarsih, 2021). This is because the assessment instrument is unable to measure students' abilities holistically.

In-depth analysis of school needs has informed the development of a research instrument that measures factual understanding and explores chemistry concepts at three levels of representation. The TTVR assessment instrument is designed to enhance learning by visualizing abstract concepts in chemistry, particularly acids and bases, using virtual reality technology.

Design and Development

The next stage is the design of the TTVR for acids and bases on the basis of the results of the analysis and the learning outcomes of the acids and base materials. The TTVR is developed by creating assessment instrument items and a storyboard that outlines the process of working on the TTVR assessment instrument for the acids and bases material.

Next, the development is carried out by assessing the feasibility of the assessment instruments and storyboards based on parameters through instrument validity (Desiriah & Setyarsih, 2021). Validity is very important, as it serves as a tool to measure the effectiveness of tests. Validity is crucial to ensure that the content aligns with the intended measurement objectives, thereby reducing bias in the research (Hartono, 2018). Content validity indicates the extent to which the research instrument accurately measures the variables intended to be measured (Heale & Twycross, 2015).

Three approaches to researching the validity of measurement tools are content validity, construct validity, and criterion validity (Suryabrata, 2005). Content validity measures the extent to which the elements of the assessment instrument and storyboard are relevant and represent the targeted construct of the measurement tool. The final outcome of content validity is an assessment of the appropriateness of the content of the assessment instrument (Ihsan, 2015). In the initial stages of instrument development, the goal of content validation is to reduce the potential variation in errors during the creation of the instrument.

Validity is assessed using a Likert scale ranging from 1 to 4. A value of V on this scale is considered good if it exceeds 0.76 (Aiken, 1985). This finding indicates that the TTVR-Etnokimia assessment instrument is suitable and relevant for measuring the researched variables. The validity of the results of the TTVR assessment instrument are presented in Table 1. The validity of the storyboard TTVR instrument material aspects is presented in Table 2, and the validity of the storyboard TTVR instrument media aspects is presented in Table 3.

Additionally, qualitative data from the validation results are analyzed to evaluate the assessment and perspectives on the TTVR-Etnokimia instrument and storyboard related to acid–base material. This analysis is based on validation feedback, including suggestions, comments, and input from experts, to enhance the development of the instrument.

TTVR-Assessment Instrument

The TTVR assessment instrument is designed to measure acid–base strength at the submicroscopic level. Its main objective is to identify misconceptions about ionization, electrical conductivity, and differences in the concentrations of H^+ and OH^- ions that affect the pH of solutions.

The development of the TTVR instrument includes four scenarios representing the stages of the keris cleansing process, using this context to illustrate the relevance of acid and base properties through the materials involved. This approach bridges scientific concepts of acids and bases with students' everyday lives.

Based on the analysis results in Table 1, item 1 was used to evaluate students' understanding of the ionization of salt solution (NaCl) in coconut water during the keris washing stage, with a high validity score of 0.90. This question focuses on the relationship between the ionization process and the ability of the salt solution to conduct electric current.

Table 2. Validity of the Content of the TTVR Assessment Instrument

Question	Value V	V Table	Conclusion
1	0.90	0.76	Valid
2	1.00	0.76	Valid
3	0.95	0.76	Valid
4	0.95	0.76	Valid
5	1.00	0.76	Valid
6	0.95	0.76	Valid
7	0.95	0.76	Valid
8	0.95	0.76	Valid
9	0.90	0.76	Valid
10	0.95	0.76	Valid
11	0.90	0.76	Valid
12	0.95	0.76	Valid
13	0.95	0.76	Valid
14	0.95	0.76	Valid
15	0.95	0.76	Valid

Items 2 to 4 assess the concept of weak acid ionization (citric acid) in lime during the keris whitening stage, with a focus on the release of ions that affect pH, with validity scores of 1.00, 0.95, and 0.95, respectively. These questions emphasize the ionization process, where ions are released in a weak acid solution, thereby affecting the pH value.

Items 5 to 10 evaluate the understanding of strong base ionization (KOH), assess students' comprehension of strong base ionization during the keris cleaning stage, with a focus on ionization reactions, pH influence, electrical conductivity, and the roles of H^+ and OH^- ions. The validity scores for these six items are very high, at 1.00, 0.95, 0.95, 0.95, 0.90, and 0.95.

Items 11 to 15 evaluate the pH testing activities and the effect of adding water on the pH of the solution. These questions focus on the influence of pH on the amounts of H^+ and OH^- ions in the solution.

The high validity results, ranging from 0.90 to 1.00, indicate that items 1–15 are effective in detecting misconceptions at the submicroscopic level, particularly related to the ionization processes of the salt solution (Santi & Rahayu, 2022), weak acid ionization, strong bases, and pH values. Thus, these factors help prevent students from merely using pH values as criteria for determining the strength of a solution and the misconception that an acidic solution is characterized by the presence of H, whereas a basic solution is characterized by the presence of OH in its chemical formula (Utami *et al*., 2020).

Therefore, the quality of the items in measuring aspects of content, language, and construct is considered valid, providing a basis for adapting the development of instruments in the context of the keris cleansing tradition. This means that, by considering these aspects, each item is designed accurately to reflect relevant acid–base material, use easily understandable language, and effectively measure students' understanding of acid–base concepts and their relation to tradition.

To improve the quality of the assessment instrument being developed, it is necessary to conduct a validity evaluation by considering suggestions and feedback from experts. Based on the experts' suggestions and feedback, several aspects of the items in the questions require improvement, including the following:

(1) Consistency in writing the items of the questions is essential. This includes aspects such as uniformity in font, punctuation, and specialized terminology when developing the TTVR-Ethnochemistry assessment instrument on acid–base topics to ensure quality. This consistency helps to avoid confusion and provides clarity for the students.

(2) The addition of information and descriptions to image elements refers to the explanation of the meaning of the image elements in each question item. The description of the image elements plays a crucial role in ensuring that students understand the visual context presented and how the images relate to the questions posed. Inaccurate descriptions can lead to subjective or even incorrect interpretations of the images, ultimately affecting students' ability to answer the questions accurately.

(3) Improving the tier-2 sentences to make them more explicit means that the TTVR instrument is constructed more clearly and is well defined. This clarity allows students to understand the expected knowledge. Therefore, the TTVR instrument on acid–base materials can significantly differentiate levels of understanding and produce more accurate assessments.

TTVR-Storyboard

The storyboard is used as a planning tool to develop research instruments integrated into virtual reality and created using Canva and Unity applications. The storyboard contains the workflow for solving problems and visualizing the questions related to the ethnochemistry of the keris cleansing process in the context of acid–base material.

The arrangement of the simulation flow in VR through the storyboard aims to present the activities that will be carried out by the students. The storyboard includes a flow that encompasses the simulation of the keris cleansing process related to acid–base materials and a simple practical simulation for determining the pH value of a solution. Through the storyboard, questions related to ethnochemistry can be systematically organized, making it easier for experts to understand as validators. The main stages of keris creation are presented in Figure 2.

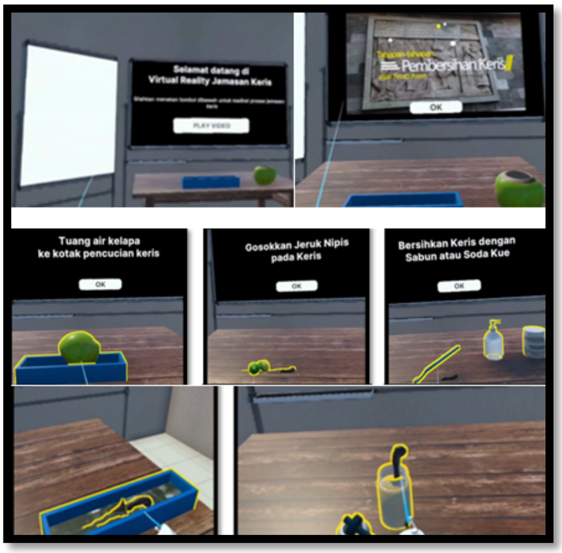

Figure 2. The main stages of jamasan keris

The storyboard flow consists of four scenarios for the process of cleansing a kris, which are integrated into virtual reality. These scenarios include washing with coconut water, whitening with lime, cleaning with soap, and a practical simulation of pH testing on baking soda. The elements within the storyboard are designed to integrate the visualization of acid–base concepts through virtual reality technology. Thus, each component can visualize three levels of representation—macroscopic, submicroscopic, and symbolic—in the study of the concepts of acids and bases.

Based on the analysis results from Tables 3 and 4, 10 indicators in terms of the media aspect and 6 indicators in terms of the material aspect of the TTVR acid–base content clearly demonstrate high validity. The high validity results analyzed using the Aiken index reveal values ranging from 0.90–1.00 for the media aspect and values ranging from 0.90–0.95 for the material aspect. These values significantly exceed the generally accepted minimum validity threshold of 0.76 (Aiken, 1985). This means that the storyboard is considered valid and suitable for use as a support tool for measuring students' conceptual understanding through TTVR.

Table 3. Validity of Storyboard Instrument TTVR material aspects

No	Material Aspect			
	Indicator	**Value V**	**V Table**	**Conclusion**
1	The suitability of the material with misconception indicators	0.76	0.90	Valid
2	The orderly organization of the material presented	0.76	0.95	Valid
3	The easy to understand the material presented	0.76	0.95	Valid
4	The subject matter presented as a question narrative is by the ethnochemistry approach	0.76	0.95	Valid
5	The images and videos can visualize material concepts	0.76	0.95	Valid
6	The systematic presentation of misconception questions is good	0.76	0.90	Valid

Table 4. Validity of Storyboard Instrument TTVR Media Aspects

No	Media Aspect			
	Indicator	**Value V**	**V Table**	**Conclusion**
1	The images in the storyboard are interesting and easy to understand	0.76	0.90	Valid
2	The sentences used are easy to understand	0.76	1.00	Valid
3	The suitability of the size of the writing in the storyboard is appropriate in each section	0.76	0.95	Valid
4	The suitability of the size of the images in the storyboard has been balanced and arranged as well as possible	0.76	0.95	Valid
5	The images and videos used in the storyboard match the theme	0.76	1.00	Valid
6	The images used in the storyboard are interesting	0.76	0.95	Valid
7	The typeface in the storyboard is easy to read	0.76	0.95	Valid
8	The combination of text and background in the storyboard is balanced and attractive	0.76	0.95	Valid
9	The color degradation used in the storyboard is appropriate	0.76	0.90	Valid
10	The images and videos already provide ethnochemistry reinforcement	0.76	0.95	Valid

Through the validity process conducted by experts, suggestions and feedback include the following:

(1) Consistency in the capitalization of letters in answer choices is crucial to prevent any option from appearing visually more prominent, which could distract students from the content. Uniformity helps learners focus on the meaning of each choice without irrelevant distractions.

(2) The addition of descriptions for the materials in the storyboard, including names and chemical formulas, helps students identify and understand the function of each material in the kris cleansing process, making it easier for them to distinguish between the shapes and characteristics of the visualized molecules. For example, providing a description of the element H in the hydrogen molecule or adding the chemical formula $C_6H_8O_7$ for citric acid.

(3) Improving the selection of image elements in the storyboard. This relates to enhancing the visualization of image elements, meaning that the chosen images must be relevant to the context of ethnochemistry during the kris era. The image elements should effectively visualize concepts in the material that are considered abstract and should not convey any other (ambiguous) meanings.

(4) The connection between the kris cleansing process and the testing simulation in the virtual laboratory needs to be bridged. This means that the storyboard in the TTVR should not only visualize the kris cleansing process but also integrate chemical principles that can be

scientifically explored through virtual laboratory simulations.

(5) Relationship between the kris cleansing process and acid–base theory. This means that the storyboard should effectively explain concepts from acid–base theory, the properties of acids and bases, and the concept of pH. Based on the suggestions and feedback from the experts, the storyboard needs to be revised. Revisions are used as a measure of improvement for the shortcomings and weaknesses of the developed product (Okpatrioka, 2023)

TTVR with previous assessment instruments

Assessment instruments are used as measurement tools to determine the extent of students' understanding, with common methods employed by teachers being essay tests and multiple-choice questions (Damayanti *et al*., 2018). These instruments are limited in their ability to measure students' cognitive ability. Conversely, assessment instruments should not only measure cognitive skills but also be designed according to the needs and characteristics of the students' learning environment, including values such as their local culture (Sukma *et al*., 2024). By integrating local culture, students will be better able to understand chemical concepts in a meaningful and contextual manner (Ode, 2024).

Conventional assessment instruments are limited in representing the abstract concepts of acids and bases, making it difficult for students to connect submicroscopic and macroscopic levels when relying solely on memorization (Malihah *et al*., 2021). This situation may lead to undetected misconceptions among students, particularly when understanding particle interactions during reactions.

Therefore, the two-tier multiple choice assessment instrument integrated with virtual reality (TTVR) becomes an effective solution. This instrument presents realistic and dynamic 3D visualizations through the kris cleansing process, such as visualization of the ionization process of citric acid ($C_6H_7O_7^-$). Unlike strong acids, which ionize completely, citric acid is a weak acid that does not fully ionize in solution. VR visualization clearly revealed that the citrate ion ($C_6H_7O_7^-$) and hydrogen ion (H^+) are separated, with H^+ bonding with H_2O to form hydronium ions (H_3O^+). Compared with the number of citric acid molecules, the amount of H+ ions produced is very small. The partial interaction between citric acid and the solvent is reversible, which is consistent with the concept of weak acid ionization, where the dominant species is nonionized acid (Chang, 2004).

Thus, the TTVR assessment instrument helps students gain a deeper understanding of acid–base concepts. Additionally, the engaging and interactive experience of using VR can enhance students' motivation and interest in learning, preventing potential misconceptions. TTVR also contributes to the development of local cultural values, making it an effective tool for creating a comprehensive and integrated learning experience. Following this, the TTVR assessment instrument will undergo a limited trial to evaluate its effectiveness in a real classroom setting. This trial involves a group of students from several schools in Surakarta.

CONCLUSION

Based on the validity results of the TTVR assessment instrument, which consists of 15 questions and 16 storyboard indicators discussed in focus group discussions (FGDs), the content validity was confirmed, with an Aiken index exceeding 0.76. This category indicates that TTVR is considered feasible; however, it is essential to implement improvements based on expert suggestions and feedback. These enhancements aim to increase the quality of the TTVR instrument, which will subsequently be tested on a limited basis in several high schools in Surakarta.

ACKNOWLEDGMENT

The authors express their gratitude to the Directorate of Research, Technology, and Community Service (DRTPM) of the Ministry of Education, Culture, Research, and Technology, as well as the Institute for Research and Community Service (LPPM) of Universitas Sebelas Maret, for supporting the Strategic Research Collaboration Program (KATALIS) under contract number 2002.1/UN27.22/ PT.01.03/2024. This support has significantly contributed to the successful implementation of this research.

REFERENCE

Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings, Educational and Psychological Measurument. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Azizah, N. L., Mahardiani, L., & Yamtinah, S. (2022). Analisis Miskonsepsi dengan Tes Diagnostik Two-Tier Multiple Choice dan In-Depth Interview pada Materi Asam Basa. *Jurnal Pendidikan Kimia FKIP Universitas Sebelas Maret*, *11*(2), 168–177. https://doi.org/10.20961/jpkim.v11i2.60345

Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The Development of a Two-Tier Multiple Choice Diagnostic Instrument for Evaluating Secondary School Students' ability to Describe and Explain Chemical Reactions using Multiple Levels of Representation. *Chemistry Education Research and Practice*, *8*(3), 293–307. https://doi.org/10.1039/B7RP90006F

Chang, R. (2004). *Kimia Dasar Jilid 2* (Edisi Ketiga). Erlangga.

Damayanti, D. R., Yamtinah, S., & Utomo, S. B. (2018). Pengembangan Instrumen Penilaiantwo-Tier Multiple Choice Question untuk Mengukur Keterampilan Proses Sains Siswa pada Tema Mata sebagai Alat Optik. *INKUIRI: Jurnal Pendidikan IPA*, *7*(2), 252. https://doi.org/https://doi.org/10.46244/numeracy.v3i2.209

Desiriah, E., & Setyarsih, W. (2021). Tinjauan Literatur Pengembangan Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi (HOTS) Fisika di SMA. *ORBITA: Jurnal Kajian, Inovasi Dan Aplikasi Pendidikan Fisika*, *7*(1), 79. https://doi.org/10.31764/orbita.v7i1.4436

Fajri, A. Y. R., Agung, S., & Saridewi, N. (2020). Penggunaan Instrumen Diagnostik Two-Tier untuk Menganalisis Miskonsepsi Asam Basa Siswa SMA dan MA. *JINoP: Jurnal Inovasi Pembelajaran*, *6*(1), 101. https://doi.org/10.22219/jinop.v6i1.8445

Gradini, E. (2016). Miskonsepsi dalam Pembelajaran Matematika Sekolah Dasar di Dataran Tinggi Gayo. *Journal of Chemical Information and Modeling*, *3*(2), 52–60. https://ejournal.bbg.ac.id/numeracy/article/view/209

Gultom, G. F., Parlindungan, J. Y., & Siregar, L. F. (2023). Analisis Miskonsepsi Peserta Didik Kelas X IPA pada Materi Ikatan Kimia Menggunakan Instrumen Two-Tier Multiple Choice Sdi SMA Negeri 1 Tanah Miring. *Arfak Chem: Chemistry Education Journal*, *6*(1), 503–515. https://doi.org/10.30862/accej.v6i1.447

Hidayatullah, A. R., Yamtinah, S., & Masykuri, M. (2022). Development of A Two-Tier Multiple-Choice Instrument Based on Higher Order Thinking Skills (HOTS) on Acids, Bases, and Salts. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 932–938. https://doi.org/10.29303/jppipa.v8i2.1423

Ihsan, H. (2015). Validitas Isi Alat Ukur Penelitian: Konsep Dan Panduan Penilaiannya. *Pedagogia*, *13*(3), 173–179.

Jamil, M. (2018). Pemanfaatan Teknologi Virtual Reality (VR) di Perpustakaan. *Buletin Perpustakaan*, *1*(1), 99–113. https://doi.org/10.1016/B978-0-12-816958-2.00007-1

Malihah, V., Farida, I., & Sari, S. (2021). Pembuatan Game Pembelajaran Chemanji Berbasis Augmented Reality (Ar) Pada Konsep Geometri Molekul Making Augmented Reality (Ar) Based Chemanji Learning Game on Molecular Geometry Concept. *Gunung Djati Conference Series*, *2*, 279–292.

Mutmainna, D., Mania, S., & Sriyanti, A. (2018). Pengembangan Instrumen Tes Diagnostik Pilihan Ganda Dua Tingkat Untuk Mengidentifikasi Pemahaman Konsep Matematika. *MaPan*, *6*(1), 56–69. https://doi.org/10.24252/mapan.2018v6n1a6

Ode, Z. (2024). Penilaian Autentik Numerasi Berbasis Kearifan Lokal pada Mata Pelajaran Matematika di Kelas V SDN 1 Kota Ternate. *Arus Jurnal Psikologi Dan Pendidikan*, *3*(2), 101–105.

Okpatrioka, O. (2023). Research And Development (R&D) Penelitian Yang Inovatif Dalam Pendidikan. *Dharma Acariya Nusantara: Jurnal Pendidikan, Bahasa dan Budaya*, *1*(1), 86–100. https://doi.org/10.47861/jdan.v1i1.154

Santi, A., & Rahayu, M. (2022). Analisis Miskonsepsi Siswa pada Materi Larutan Elektrolit dan Non Elektrolit Menggunakan Instrumen Multirepresentasi Four-Tier Diagnostic Test Berbasis Piktorial. *Ournal of Chemical Education*, *11*(3), 210–219. https://doi.org/10.26740/ujced.v11n3.p210-219

Saputri, L., Dewi Muldayanti, N., & Eka Setiadi, A. (2016). Analisis Miskonsepsi Siswa Dengan Certainty Response Index (CRI) Pada Submateri Sistem Saraf Dikelas XI IPA SMA Negeri 1 Selimbau. *Jurnal Bioeducation*, *3*(2). http://dx.doi.org/10.29406/186

Setyoko, H., Mulyani, S., & Yamtinah, S. (2018). The Implementation of Problem-Solving Model Using Concept Mapping Strategy to Increase Students' Interest and Learning Achievement at the Lintas-Minat Chemistry Class. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *2*(3), 178. https://doi.org/10.20961/jkpk.v2i3.16780

Siswaningsih, W., Firman, H., & Rofifah, R. (2015). Pengembangan Tes Diagnostik Two-Tier Berbasis Piktorial Untuk Mengidentifikasi Miskonsepsi Siswa Pada Materi Larutan Elektrolit Dan Nonelektrolit. *Jurnal Pengajaran Matematika Dan Ilmu Pengetahuan Alam*, *20*(2), 144. https://doi.org/10.18269/jpmipa.v20i2.577

Sokpheng, S., & Meng, L. H. (2022). Comparison of Modern and Conventional Learning Methods for Children with Special Needs. *Journal of Asian Multicultural Research for Educational Study*, *3*(2), 14–22.

Sudaryono, Rahardja, U., Aini, Q., Isma Graha, Y., & Lutfiani, N. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, *1364*(1). https://doi.org/10.1088/1742-6596/1364/1/012050

Sugiyono. (2016). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*.

Sukma, S., Asrianti, A., Efendi, P., & Lembah, G. (2024). *Pengembangan Instrumen Penilaian dalam Pembelajaran Membaca Bermuatan Etika Budaya di SMP Kota Palu*. *10*(4), 4606–4613.

Suryabrata, S. (2005). *Pengembangan Alat Ukur Psikologis*. Yogyakarta: Penerbit Andi.

Susanto, F., & Ayuni, I. (2017). Pengembangan Perangkat Pembelajaran Model Kooperatif Tipe NHT dengan Strategi Pemecahan Masalah (Problem Solving) Sistematis bagi Peserta Didik SMP di Kabupaten Pringsewu. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *6*(3), 301. https://doi.org/10.24127/ajpm.v6i3.1054

Utami, I., Mulyani, B., & Yamtinah, S. (2020). Identifikasi Miskonsepsi Asam-Basa dengan Two Tier Multiple Choice dilengkapi Interview. *Jurnal Pendidikan Kimia*, *9*(1), 89–97. https://doi.org/https://doi.org/10.20961/jpkim.v9i1.34078

Wartoni, & Benyamin, P. I. (2020). Strategi Pengembangan Tes Objektif (Pilihan Ganda). *Diegesis : Jurnal Teologi*, *5*(1), 1–8. https://doi.org/10.46933/DGS.vol5i1%25p