# Identification Test of Ability to Understand Multiple Representation in Basic Chemistry I Course: Validity and Reliability

Syahrial Syahrial*, Sri Winarni
*Chemistry Education Department, Syiah Kuala University*
*Corresponding Author**

***Abstract*: The study objective was to determine the validity and reliability of the test items used to measure understanding of multiple representations. For this purpose, quantitative methods are applied. Participants were first-year students of the Chemistry Department, Education and Teacher Training Faculty, Syiah Kuala University, who took the Basic Chemistry course I (specifically solubility, redox, and hydrocarbons). The test was followed voluntarily. Before determining the validity and reliability of each test, the Multiple Representation Understanding Test (MRUT) was developed, which was conducted in five stages. MRUT contains 20 items, and its validity is determined using Pearson Product Moment (PPM). Valid test items are nine where $r_{count}$t 0.3128-0.7145. The nine-item tests are reliable, and Cronbach alpha ranged from 0.701 to 0.769 (moderate-high).**

***Keywords*: multiple-choice, reliability, test, validity, multiple-representation**

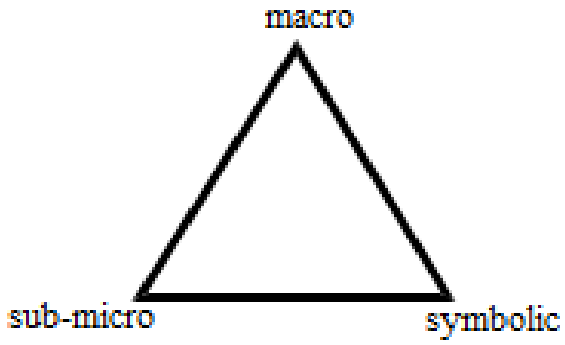
## I. INTRODUCTION

The idea that all matter in nature is particulate (Gilbert and Treagust 2009) made us chemistry teachers aware of the need to introduce students to micro-level understanding skills as early as possible. The ability to understand the micro-level is needed in order to be able to understand the explanation of various phenomena in chemistry (Russell and Kozma 1997; Wang et al. 2014), such as atomic models, how atoms bond to form molecules, explain differences in the state of substances due to temperature changes, and many other phenomena. Many studies have shown that the ability to understand at the micro-level helps students succeed in chemistry studies.

Many students have difficulty understanding the micro-level in learning chemistry (Gabel 1993, 1999). Even though this chemical ability is an understanding of chemical concepts (Chandrasegaran et al. 2008; Sanger 2005) and can be a predictor of student success in learning chemistry (Cheng and Gilbert 2009; Devetak et al. 2007). Thankfully, students' micro-level understanding can be improved through appropriate practice or learning, using various types of representation (Mcdermott and Hand 2013).

A number of studies have found that teaching abstract concepts that require micro-understanding can be conducted multiple representations (Ainsworth 1999; Sim et al. 2014), analogies (Çalik and Ayas 2005; Özmen and Kenan 2007), and various types of mental models (Coll 2008). Multiple representations in this context are multiple external representations that are defined as expressing something or a phenomenon in various forms of expressions such as graphs, pictures, tables, schemes, sketches, and symbols (Hinton and Nakhleh 1999). An *analogy* is defined as correspondence in several ways between different concepts, principles, or formulas (Thiele and Treagust 1991). An analogy is a mapping between the same features of the concept, principle, and formula. Meanwhile, the mental model is defined as a mental image form representing personal mental constructions (Johnson-Laird 1980) or shows a person's belief in a system (Gentner and Stevens 1983).

Multiple representations widely used in chemistry studies are the triple chemistry of Johnstone (Johnstone 1991). Johnstone described the chemical triple as an interrelationship between macro, symbolic, and micro. This reciprocal relationship is known as the Johnstone triangle.




Gambar 1: Johnstone triangle

Johnstone explained that the macro representation shows everything that the senses can feel, namely heard, smelled, seen, felt, or tasted (it is important to remember that not all chemicals can be tasted). For example, when students dissolve table salt, NaCl. Students will see clearly how the solid crystalline NaCl will be lost in the water solvent, and the water solvent does not change color except taste. Symbolic is a representation as a symbol, letter, or number (at this point it has been interpreted as a representation only, pictures, graphs, tables, and so on). Sub-microscopic representation describes a

molecular model whose shape, size, and color conform to the scientific agreement.

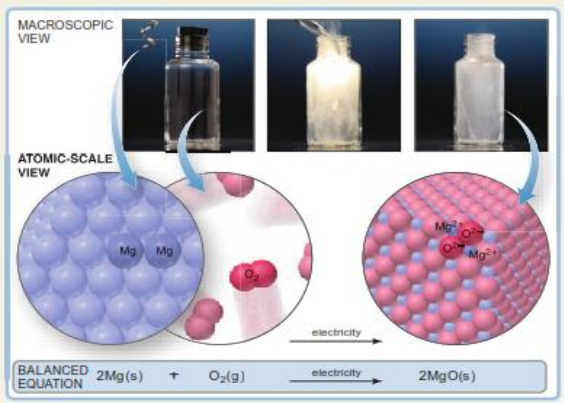



Figure 1: Reactionof magnesium withoxygen(Silberberg 2009)

After implementing multiple representation-based learning, it is necessary to develop instruments that can measure student understanding. The instrument developed must be based on multiple representations. If a general instrument is used, the students' real understanding will not be measured, especially the students' microscopic level understanding. Therefore, in Basic Chemistry I course in the Chemistry Education Department, it is necessary to develop exam questions following the multiple representation-based lecture process. Students are asked to analyze molecular images, graphs, and tables.

The assessment is in multiple-choice questions with four answer choices consisting of one correct answer and three distractors. The questions developed need to be tested for accuracy (validation) (Messick 1989, 1995; Rovinelli and Hambleton 1977) and repeatability (reliability) (Jonsson and Svingby 2007) so that they can be used to measure students' level of multiple representations understanding

## II. METHODOLOGY

Because the study aims to determine the validity and reliability of the test items, this study uses quantitative methods. Thus in data analysis using statistics. The following are described the study methodology.

*Test Instrument Development*

The test instrument is developed through the following stages. (1) Analysis of the description and content in the Basic Chemistry I course in theChemistry Education Department, Education and Teacher Training Faculty, Syiah Kuala University; (2) analysis of the assessment instruments used previously in the Basic Chemistry I course (specifically the topic of solubility, redox, and hydrocarbons); (3) analysis of the content and questions in several general chemistries and introductory chemistry textbooks published in the last ten years; (4) examining various sample questions from journal articles; and (5) development of 20 item tests multiple-choice form assessment instruments.

*Participants*

Participants in this study were first-year students of the Chemistry Education Department, Education and Teacher Training Faculty, Syiah Kuala University, who took the Basic Chemistry I course (specifically the topic of solubility, redox, and hydrocarbons). The number of participants is 79 people who are dominated by women (only five men) and voluntarily.

*Procedure*

Data collection was carried out by giving written tests to 79 participants at the same time. Participants were given 60 minutes to complete a set of multiple-choice questions in the form, i.e., the Multiple Representation Understanding Test (MRUT), which contained 20 items. The sitting position of the participants is designed so that they are not close to each other. A longer time was given because the participants had never had the experience to solve multiple representation-based questions.

*Data Analysis*

Construct validity determination of the test instrument using Pearson Product Moment (PPM) (Yandriani et al. 2020). PPM use is the most popular and widely used procedure for researchers to determine the instrument's validity (Mehrens and Lehmann 1991). The test item is valid if the PPM correlation coefficient, $r > 0.3$ (Pallant 2011; Yandriani et al. 2020).

Valid test items are determined for reliability using Cronbach's alpha (Mohamad et al., 2015; Sadhu and Laksono 2018; Taber 2018). The Cronbach alpha as an instrument's quality indicator was due to its high use by researchers (Taber 2018), except for ordinal data, which proved to be less sensitive (Zumbo et al. 2007). All data analysis was carried out with IBM SPSS statistical software version 16 help. Meanwhile, guidelines for determining the level of reliability were used as in Table 1 (Gottems et al. 2018).

Table 1. Cronbach Alpha Interpretation

| Cronbach Alpha (α) | Conclusion |
|---|---|
| ≤ 0.30 | Very Low |
| 0.30 <α ≤ 0.60 | Low |
| 0.60 <α ≤ 0.75 | Moderate |
| 0.75 <α ≤ 0.90 | High |
| α> 0.90 | Very high |

## III. RESULTS AND DISCUSSION

*Test Instrument Development*

The results of assessment instruments analysis previously used in the Basic Chemistry I course (specifically the topic of solubility, redox, and hydrocarbons) and a review of 3 general chemistry or basic chemistry books resulted in some findings that became the basis for MRUT development. Five books

were reviewed on the topic of solubility, redox, and hydrocarbons, namely.

1. Chemistry: the molecular nature of matter and change (2009) written by Martin S. Silberberg;
2. Chemistry & Chemical Reactivity (2019) written by John C. Kotz et al.; and
3. Chemistry: an atoms-focused approach (2018) written by Thomas R. Gilbert et al.
4. Meanwhile, 3 articles were reviewed, namely:
5. Connection making between multipleple graphical representations: A multiple-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry (2105) written by Rau et al;
6. Student-generated submicro diagrams: a useful tool for teaching and learning chemical equations and stoichiometry (2010) written by Davidowitz et al; and
7. Examination of Secondary School Students' Ability to Transform among Chemistry Representation Levels Related to Stoichiometry (2020) written by Çelikkıran. The interesting findings from the three articles i.e: (1) many students do not yet understand multiple representations; and (2) it turns out that the algorithmic test items can be presented in sub-microscopic views.

The exciting findings from the three articles, i.e.:

1. Many students do not yet understand multiple representations; and
2. It turns out that the algorithmic test items can be presented in sub-microscopic views.

In the final stage, MRUT developmental is carried out. As consideration for the development of each test item is all findings and Chemistry Education Department curriculum. The development results are 20 test items with the characteristics as presented in Table 2.

Table 2. Characteristics of Test Items Topics Solubility, Redox, and Hydrocarbons

| Jenis Tes | Characteristics |
|---|---|
| MRUT | Involves multiple representations, predominantly conceptual, case (daily case), HOTS |
| Conventional | Factual, predominantly algorithmic, not all HOTS |

MRUT was developed to familiarize students with understanding multiple representations, thinking critically, and understanding problems through solving cases, especially those that are daily cases. This competence is a demand for `21st-century learning (Tight 2020) and the 4.0 era (Liliasari 2018)

*Construct Validity*

An instrument will produce a good measurement if construct validity is fulfilled (Mohajan 2017). Construct validity indicates the extent to which specific measures are consistently related to other measures regarding the concept being measured (Nunnally and Bernstein 1994; Thatcher 2010). Instruments that do not meet the criteria for construct validity cannot be used in the measurement. Therefore, 6 out of 20 MRUT items had to be discarded because they were invalid. The discarded items were numbers 2, 5, 6, 8, 10, 12, 13, 14, 15, 16, 18, and 19.

Table 3 shows the Pearson product-moment correlation coefficient, the *r* test item used to identify the ability to understand multiple representations.

Table3. MRUT Item Coefficient, *r*

| No. Item Tes | $r_{calculated}$ | Conclusion |
|---|---|---|
| 1 | 0.5319 | Valid |
| 2 | -0.0763 | Invalid |
| 3 | 0.4951 | Valid |
| 4 | 0.7145 | Valid |
| 5 | 0.2733 | Invalid |
| 6 | -0.0674 | Invalid |
| 7 | 0.3176 | Valid |
| 8 | 0.1152 | Invalid |

Table3. MRUT Item Coefficient, *r*

| No. Item Tes | $r_{calculated}$ | Conclusion |
|---|---|---|
| 9 | 0.4871 | Valid |
| 10 | 0.2226 | Invalid |
| 11 | 0.3128 | Valid |
| 12 | 0.2145 | Invalid |
| 13 | 0.5722 | Valid |
| 14 | 0.2582 | Invalid |
| 15 | -0.0951 | Invalid |
| 16 | -0.2080 | Invalid |
| 17 | 0.4819 | Valid |
| 18 | 0.2344 | Invalid |
| 19 | 0.2341 | Invalid |
| 20 | 0.5016 | Valid |

After undergoing measurement, only 9 or 45% of the remaining MRUT items remained. Test item reduction due to invalid, and it's possible, occurs in several tests developed previously. Several tests whose items were reduced by up to 55% were reported by Yazar and Nakiboğlu (2019), 5.6% by Setyawaty et al. (2018), 45.5% (Chandrasegaran et al. 2007), and 66.7% by Anil et al. (2010). This fact indicates that evaluation instruments development needs to be carried out carefully and requires a thorough, comprehensive study of the concepts to be identified.

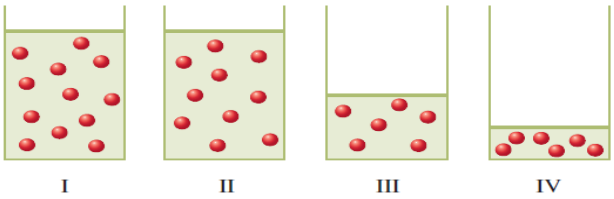
The multiple-choice test construction does not only focus on a stem, stimulant, and correct answers and is mono-

interpretation, but also needs to focus on distractors. According to Tarrant and Ware (2010), even the number of distractors can affect the performance of multiple-choice tests. A good test item must meet some standards, including truly representing the domain being tested, undergoing cross-validation, and only being sensitive to the characteristics of the concept being tested and others (Wise and Plake 2016).

.On the other hand, a test item may not be constructively valid but valid in terms of content, especially if there is nothing conceptually wrong. According to Ackerman (1991), the cause of invalid test items is too many abilities as measured by one test item. Another cause is the difference in understanding ability or sensitivity to tests from different test groups. This phenomenon overshadows the test items number 2, 5, 6, 8, 10, 12, 14, 15, 16, 18, and 19. The eleven test items are full of prerequisite concepts so that their sensitivity is very different from the test group—for example, the following test item number 12.

MRUT Item number 12:

*"The diagram below represents the same solute in a varying number of concentrations.*




*Two diagrams that represent two solution systems that have the same concentration are .... (answer: A)*

*A. I and III C. II and IV*
*B. III and IV D. I and II*

Item number 12 can be answered if students understand:

1. A molecular diagram is a direct representation of substances amount in the system.
2. Two systems are stated to have the same concentration if the substance amount and the amount of solvent in the two systems are the same.
3. In both systems, the ratio between the solute and the solvent is the same.

Another cause of an invalid test item, according to Haladyna is the participant's insincere response and cheating (Haladyna 2004). This MRUT is a paper and pencil test. Participants should finish MRUT in 60 minutes. The provision of long duration due to the test type and multiple representations used in lectures is new for the participants. The test items presented with the new model can cause participants to be disinterested or feel uncomfortable. Two such conditions also can cause the test item to be invalid (Haladyna and Rodriguez 2013)).

*Reliability*

The reliability of the test items is only aimed at valid test items (Kara and Çelikler 2015; Tarrant and Ware 2012). The result of the calculation shows Cronbach Alpha calculation result is 0.761 at $p = 0.05$ ($r$table = 0.666 at $p = 0.005$). Thus, MRUT can be classified as having high internal consistency. The instrument used to measure or identify distinctive abilities has high internal consistency (Cronbach alpha, α) as reported by Cooper and Sandi-Urena (2009) 0.87 at pretest and 0.91 at post-test; Zapata-Caceres et al. (2020) 0.824; and Yandriani et al. (2020).

Internal consistency shows the degree to which all items measure the same concepts or constructs of people who take the test and shows the degree of items freedom from measurement errors (Tavakol and Dennick 2011; Thorndike and Thorndike-Christ 2014). Furthermore, according to Tavakol and Dennick (2011), Cronbach alpha quantitatively the internal consistency. For example, if the MRUT has Cronbach alpha = 0.761, then the variance error is (0.761 x 0.761 = 0.579; 1 - 0.579 = 0.421). Error variance 0.421 closest to 42.1%, there is a measurement error. It should be remembered that a high Cronbach alpha does not necessarily mean that internal consistency is high because other factors come into play, namely the long test. The longer the test (the number of test items is large), the higher the internal consistency (Kline 1994; Streiner 2003). Furthermore, Tavakol and Dennick (2011) state that instrument Cronbach alpha is measured based on the score from a particular sample, so it is necessary to repeat the measurement if the instrument is used on a different sample.

Although various literature states that the Cronbach alpha limit for a reliable test item varies widely, according to Murphy and Davidshofer (2005), it depends on the purpose of using the test. For example, Cronbach alpha is 0.70 for initial research, 0.8 for tools in basic research, and 0.90 for clinical research (Streiner 2003). According to Kane (1986), the Cronbach alpha minimum value for a test item is said to be reliable is 0.5. Nunnally and Bernstein (1994) state that the test item Cronbach alpha ≤ 0.6 can be considered a flawed measure in a different place. Table 4. shows the Cronbach alpha value for each MRUTitems..Tabel 4**.** Cronbach alpha Coefficient for Items MRUT

Tabel 4.Cronbach alpha Coefficient for Items MRUT

| No. Test Item | Cronbachalpa | Reliability |
|---|---|---|
| 1 | 0.701 | Moderate |
| 3 | 0.735 | Moderate |
| 4 | 0.753 | High |
| 7 | 0.755 | High |
| 9 | 0.741 | Moderate |
| 11 | 0.769 | High |
| 13 | 0.747 | Moderate |
| 17 | 0.717 | Moderate |
| 20 | 0.724 | Moderate |

Cronbach alpha for all MRUT items ranged from 0.701-0.769. Referring to Gottems et al. (2018) and Kane (1986), all valid

test items are reliable to measure the ability of students to understand several representations in Basic Chemistry I lectures. Cronbach alpha obtained is almost the same as the reliability of the Scientific Reasoning Test items (SRT), namely 0.71 for pretest, 0.61 for post-test, and 0.76 for retention-test (Lee and She 2010). This MRUT item turned out to have a higher level of reliability when compared to the first tier for the three-tier test developed by Caleon and Subramaniam (2010), namely 0.58 on the first level test and 0.63 on the second level test and also multiple level representation tests, 0.65 (Chandrasegaran, at al. 2008) and the Simple Electric Circuit Diagnostic Test (SECDT) (Peşman and Eryilmaz 2010), namely 0.69.

## IV. CONCLUSIONS

Learning success can be identified through assessment or evaluation. However, the instruments used will determine whether the implementation of the assessment or evaluation is carried out correctly or not. Therefore, a valid and reliable instrument is needed to arrive at a precise conclusion of an assessment or evaluation process.

Learning involves several representations certainly requires instruments according to assessment or evaluation. For this purpose, it is essential to develop MRUT consisting of 20 items. MRUT is used to identify student's ability to understand multiple representations in the Basic Chemistry I course (specifically the topic of solubility, redox, and hydrocarbons). The assessment or evaluation instrument requirements are feasible to use; it meets minimum validity and reliability standards. After MRUT analysis, it turns out that only 9 out of 20 MRUT items were suitable for use.

It is suspected that the reason for test item inadequacy is that it contains too many prerequisite concepts. Therefore, it is suggested for teachers to develop test items that contain not too many prerequisite concepts. Another suggestion is also to study the quality of distractor, option, and stem construction of a multiple-choice test.

## REFERENCES


[1] Ackerman, T. A. (1991) 'A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multipledimensional Perspective,' in Annual Meeting of the American Educational Research Association.

[2] Ainsworth, S. (1999) 'The functions of multiple representations,' Computers and Education, 33(2–3), pp. 131–152. DOI: 10.1016/s0360-1315(99)00029-9.

[3] Anil, D. et al. (2010) 'Level determination exam (SBS-2008) the determination of the validity and reliability of 7th grade mathematics sub- test', Procedia - Social and Behavioral Sciences, 2(2), pp. 5292–5298. DOI: 10.1016/j.sbspro.2010.03.863.

[4] Caleon, I. and Subramaniam, R. (2010) 'Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves,' International Journal of Science Education. DOI: 10.1080/09500690902890130.

[5] Çalik, M. and Ayas, A. (2005) 'An analogy activity for incorporating students' conceptions of types of solutions,' Asia-Pacific Forum on Science Learning and Teaching, 6(2), pp. 1–13.

[6] Chandrasegaran, A. L., Treagust, D. F. and Mocerino, M. (2007) 'The development of a two-tier multiple ply-choice diagnostic instruments for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation,' Chemistry Education Research and Practice, 8(3), pp. 293–307. DOI: 10.1039/B7RP90006F.

[7] Chandrasegaran, A. L., Treagust, D. F. and Mocerino, M. (2008a) 'An evaluation of a teaching intervention to promote students' ability to use multiple levels of representation when describing and explaining chemical reactions,' Research in Science Education, 38(2), pp. 237–248. DOI: 10.1007/s11165-007-9046-9.

[8] Chandrasegaran, A. L., Treagust, D. F. and Mocerino, M. (2008b) 'An Evaluation of a Teaching Intervention to Promote Students' Ability to Use Multipleple Levels of Representation When Describing and Explaining Chemical Reactions,' Research in Science Education, 38(2), pp. 237–248. DOI: 10.1007/s11165-007-9046-9.

[9] Cheng, M. and Gilbert, J. K. (2009) 'Towards a Better Utilization of Diagrams in Research into the Use of Representative Levels in Chemical Education,' in Gilbert, J. K. and Treagus, D. (eds) Models and Modeling in Science Education Multipleple Representations in Chemical Education John. Amsterdam: Springer-Science, p. 369.

[10] Coll, R. K. (2008) 'Chemistry Learners' Preferred Mental Models for Chemical Bonding,' Journal Of Turkish Science Education, 5(1), pp. 22

[11] Cooper, M. M., and Sandi-Urena, S. (2009) 'Design and validation of an instrument to assess metacognitive skillfulness in chemistry problem solving, Journal of Chemical Education, 86(2), pp. 240–245. DOI: 10.1021/ed086p240.

[12] Devetak, I., Vogrinc, J. and Glažar, S. A. (2007) 'Assessing 16-year-old students' understanding of Aqueous solution at submicroscopic level', Research in Science Education, 39(2), pp. 157–179. DOI: 10.1007/s11165-007-9077-2.

[13] Gabel, D. (1999) 'Improving Teaching and Learning through Chemistry Education Research: A Look to the Future, Journal of Chemical Education, 76(4), p. 548. DOI: 10.1021/ed076p5

[14] Gabel, D. L. (1993) 'Use of the particle nature of matter in developing conceptual understanding,' Journal of Chemical Education, 70(3), pp. 193–194. DOI: 10.1021/ed070p193.

[15] Gentner, D. and Stevens, A. L. (1983) Mental Models. New Orleans: Lawrence Erlbaum Associates, Inc. doi:10.1017/CBO9781107415324.004.

[16] Gilbert, J. K., & Treagust, D. F. (2009) 'Introduction: Macro, Submicro and Symbolic Representations and the Relationship Between Them: Key Models in Chemical Education,' in Gilbert, J. K., and Treagust, D. F. (eds) Multipleple Representations in Chemical Education. Dodrecht: Springer Science+Business Media B.V, p. 367. DOI: 10.1017/CBO9781107415324.004.

[17] Gottems, L. B. D. et al. (2018) 'Good practices in normal childbirth: Reliability analysis of an instrument by cronbach's alpha,' Revista Latino-Americana de Enfermagem, 26. DOI: 10.1590/1518-8345.2234.3000.

[18] Haladyna, T. M. (2004) Developing and Validating Multipleple-Choice Test Items. Mahwah New Jersey: Lawrence Erlbaum Associates, Publishers. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

[19] Haladyna, T. M. and Rodriguez, M. C. (2013) Developing and validating test items. New Orleans: Routledge. DOI: 10.4324/9780203850381.

[20] Hinton, M. E., and Nakhleh, M. B. (1999) 'Students' Microscopic, Macroscopic, and Symbolic Representations of Chemical Reactions,' The Chemical Educator, 4(5), pp. 158–167. DOI: 10.1007/s00897990325a.

[21] Johnson-Laird, P. N. (1980) 'Mental models in cognitive science,' Cognitive Science, 4(1), pp. 71–115. DOI: 10.1016/S0364-0213(81)80005-5.

[22] Johnstone, A. H. (1991) 'Why is science difficult to learn? Things are Seldom What They Seem', Journal of Computer Assisted Learning, 7, pp. 75–83.

[23] Jonsson, A. and Svingby, G. (2007) 'The use of scoring rubrics: Reliability, validity and educational consequences,' Educational Research Review, 2(2), pp. 130–144. DOI: 10.1016/j.edurev.2007.05.002.

[24] Kane, M. T. (1986) 'The Role of Reliability in Criterion-Referenced Tests', Journal of Educational Measurement, 23(3), pp. 221–224.
[25] Kara, F. and Çelikler, D. (2015) 'Development of Achievement Test : Validity and Reliability Study for Achievement Test on Matter Changing,' Journal of Education and Practice, 6(24), pp. 21–27.
[26] Kline, P. (1994) An Easy Guide to Factor Analysis, An Easy Guide to Factor Analysis. London: Routledge. DOI: 10.4324/9781315788135.
[27] Lee, C. Q. and She, H. C. (2010) 'Facilitating students' conceptual change and scientific reasoning involving the unit of combustion,' Research in Science Education, 40(4), pp. 479–504. DOI: 10.1007/s11165-009-9130-4.
[28] Liliasari, S. (2018) 'Modeling skill in chemistry education to win students on global competitiveness,' in Rahmawati, Y. and Taylor, P. C. (eds) Empowering Science and Mathematics for Global Competitiveness. Leiden: CRC Press.
[29] Mcdermott, M. A., and Hand, B. (2013) 'The impact of embedding multiple modes of representation within writing tasks on high school students' chemistry understanding,' Instructional Science, 41, pp. 217–246. DOI: 10.1007/s11251-012-9225-6.
[30] Mehrens, W. A. and Lehmann, I. J. (1991) Measurement and Evaluation In Education and Psychology. Wadsworth, Belmont: Holt, Rinehart, and Winston.
[31] Messick, S. (1989) 'Meaning and Values in Test Validation : The Science and Ethics of Assessment,' Educational Researcher, 18(5). DOI: 10.3102/0013189X018002005.
[32] Messick, S. (1995) 'Standards of Validity and the Validity of Standards in Performance Assessment,' Educational measurement: Issues and practice,
[33] Mohajan, H. K. (2017) 'Two Criteria for Good Measurements in Research: Validity and Reliability, Annals of Spiru Haret University. Economic Series, 17(4), pp. 59–82. DOI: 10.26458/1746.
[34] Mohamad, M. M. et al. (2015) 'Measuring the Validity and Reliability of Research Instruments', Procedia - Social and Behavioral Sciences, 204(November 2014), pp. 164–171. DOI: 10.1016/j.sbspro.2015.08.129.
[35] Murphy, K. R. and Davidshofer, C. O. (2005) Psychological Testing Prine I pies and Applications Principles of Psychological Measurement. New Jersey: Upper Sadler River.
[36] Nunnally, Jum C, and Bernstein, I. H. (1994) Psychometric Theory. New York: McGraw-Hili, Inc. DOI: 34567890 DOCmoC 998765 ISBN.
[37] Nunnally, Jum C., and Bernstein, I. H. (1994) Psychometric Theory. New York: McGraw Hill. Available at: https://books.google.com/books?id=_6R_f3G58JsC&pgis=1.
[38] Özmen, H. and Kenan, O. (2007) 'Determination of the Turkish primary students' views about the particulate nature of matter,' Asia-Pacific Forum on Science Learning and Teaching, 8(1), pp. 1–15.
[39] Pallant, J. (2011) SPSS Survival Manual website. New York: Open University Press.
[40] Peşman, H. and Eryilmaz, A. (2010) Development of a three-tier test to assess misconceptions about simple electric circuits, Journal of Educational Research, 103(3), pp. 208–222. DOI: 10.1080/00220670903383002.
[41] Rovinelli, R. J. and Hambleton, R. K. (1977) 'On the Use of Content Specialists in the Assessment of Criterion-Referenced Test Item Validity,' Dutch Journal of Educational Research, 2(49–60), pp. 49–60.
[42] Russell, J. W., and Kozma, R. B. (1997) 'Use of Simultaneous-Synchronized Macroscopic , Microscopic , and Symbolic Representations To Enhance the Teaching and Learning of Chemical Concepts,' Journal of Chemical Education, 74(3), pp. 330–334.
[43] Sadhu, S. and Laksono, E. W. (2018) 'Development and validation of an integrated assessment for measuring critical thinking and chemical literacy in chemical equilibrium,' International Journal of Instruction, 11(3), pp. 557–572. DOI: 10.12973/iji.2018.11338a.
[44] Sanger, M. J. (2005) 'Evaluating students' conceptual understanding of balanced equations and stoichiometric ratios using a particulate drawing,' Journal of Chemical Education, 82(1), pp. 131–134. DOI: 10.1021/ed082p131.
[45] Setyawaty, R. et al. (2018) 'Validity Test and Reliability of Indonesian Language Multipleple Choice in Final Term Examination,' KnE Social Sciences, 3(9), p. 43. DOI: 10.18502/kss.v3i9.2609.
[46] Silberberg, M. S. (2009) Chemistry: the molecular nature of matter and change. Fifth, Chemistry. Fifth. New York: McGraw Hill.
[47] Sim, J. H., Gnanamalar, E., and Daniel, S. (2014) 'Representational competence in chemistry : A comparison between students with different levels of understanding of basic chemical concepts and chemical representations A comparison between students with different,' pp. 1–17. DOI: 10.1080/2331186X.2014.991180.
[48] Streiner, D. L. (2003) 'Starting at the beginning: An introduction to coefficient alpha and internal consistency, Journal of Personality Assessment, 80(1), pp. 99–103
[49] Taber, K. S. (2018) 'The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education, Research in Science Education, 48(6), pp. 1273–1296. DOI: 10.1007/s11165-016-9602-2.
[50] Tarrant, M. and Ware, J. (2010) 'A comparison of the psychometric properties of three-and four-option multipleple-choice questions in nursing assessments,' Nurse Education Today, 30(6), pp. 539–543.
[51] Tarrant, M. and Ware, J. (2012) 'A Framework for Improving the Quality of Multipleple-Choice Assessments', Nurse Educator, 37(3), pp. 98–104. DOI: 10.1097/NNE.0b013e31825041d0.
[52] Tavakol, M. and Dennick, R. (2011) 'Making sense of Cronbach's alpha,' International journal of medical education, 2, pp. 53–55. DOI: 10.5116/ijme.4dfb.8dfd.
[53] Thatcher, R. W. (2010) 'Validity and reliability of quantitative electroencephalography,' Journal of Neurotherapy, 14(2), pp. 122–152. DOI: 10.1080/10874201003773500.
[54] Thiele, R. B., and Treagust, D. F. (1991) Using Analogies To Aid Understanding in Secondary Chemistry Education. Perth.
[55] Thorndike, R. M. and Thorndike-Christ, T. (2014) Measurement and Evaluation in Psychology and Education. Edinburgh: Pearson Education Limited. DOI: 10.2307/2282039.
[56] Tight, M. (2020) 'Twenty-first century skills: meaning, usage and value,' European Journal of Higher Education, 9. DOI: 10.1080/21568235.2020.1835517.
[57] Wang, Z. et al. (2014) 'Chemistry Teachers' Knowledge and Application of Models,' J Sci Educ Technol, 23, pp. 211–226. DOI: 10.1007/s10956-013-9455-7.
[58] Wise, L. L., and Plake, B. S. (2016) 'Test Design and Development Following the Standards for Educational and Psychological Te sting,' in Suzanne Lane, Raymond, M. R., and Haladyna, T. M. (eds) Handbook of Test Development. New York: Routledge. DOI: 10.4324/9780203102961.
[59] Yandriani, Rery, R. U. and Erna, M. (2020) 'Validity and Reliability of Assessment Instruments for Analytical Thinking Properties'Ability and Chemical Literacy in the Colligative, Journal of Physics: Conference Series, 1655(1). DOI: 10.1088/1742-6596/1655/1/012056.
[60] Yazar, O. G., and Nakiboğlu, C. (2019) 'Development of Achievement Test about Unit of "Nature and Chemistry" for 9th Grades: A Validity and Reliability Study', 13(1), pp. 76–104. DOI: 10.17522/balikesirnef.571399.
[61] Zapata-Caceres, M., Martin-Barroso, E. and Roman-Gonzalez, M. (2020) 'Computational thinking test for beginners: Design and content validation,' IEEE Global Engineering Education Conference, EDUCON, 2020-April, pp. 1905–1914. DOI: 10.1109/EDUCON45650.2020.9125368.
[62] Zumbo, B. D., Gadermann, A. M. and Zeisser, C. (2007) 'Ordinal versions of coefficients alpha and theta for likert rating scales,' Journal of Modern Applied Statistical Methods, 6(1), pp. 21–29. DOI: 10.22237/jmasm/1177992180.