Validation of Multiple Representation Instrument to Measure Student’s Multiple Representation Skill

Vika Puji Cahyani1,*, Hari Sutrisno1,2

1Chemistry Education Masters and Doctoral Programs, Graduate School, Yogyakarta State University, Yogyakarta 55281, Indonesia
2Department of Chemistry Education, Faculty of Mathematics and Natural Sciences, Yogyakarta State University, Yogyakarta 55281, Indonesia
*Corresponding author: vikapujicahyani@yahoo.co.id

Received July 03, 2018; Revised August 12, 2018; Accepted August 23, 2018

Abstract This research aims to validate multiple representations instruments for measuring student’s multiple representations skill of electrolytes and nonelectrolytes. This research uses quantitative descriptive method with nonexperimental approach. This research applies stratified sampling method with total 94 students who respond to 40 items multiple choices that cover macroscopic, microscopic, symbolic, and mathematical aspects. Rasch model in this research analyzes the instruments such as determining the unidimensionality, reliability, item map, item measure, item fit, and test information function. Data analysis shows that the finding of good category of validity and reliability of the instrument can be used as a diagnostic tool in the assessment of student’s multiple representation abilities.

Keywords: descriptive quantitative, multiple representation, rasch model, electrolyte and nonelectrolyte, psychometric properties

Cite This Article: Vika Puji Cahyani, and Hari Sutrisno, “Validation of Multiple Representation Instrument to Measure Student’s Multiple Representation Skill.” American Journal of Educational Research, vol. 6, no. 8 (2018): 1198-1205. doi: 10.12691/education-6-8-20.

1. Introduction

Chemistry is the study of the structure, properties, and reactions of element and substance. Chemistry covers the understanding of macroscopic, submicroscopic, and symbolic levels [1]. The use of different representation can help learners connect one level to another in a better way. Conceptual understanding is the key aspect of chemistry learning. Learners can have strong chemical concepts understanding when they can connect their insights in different representations [2,3,4] and how to relate each new concept or fact in the three domains that is macroscopic: how chemical phenomena can be observed by using the five senses of human such as color, smell etc.). Submicroscopic is the interaction or the form of invisible molecules including atoms, molecules, ions, and so on. Symbolic is the representation that consists of formulas, equations, symbols, mathematics, and graphics [5,6]. However, chemistry teaching rarely helps students connecting some representations. This teaching method often leads learners to confusion which has negative consequences on students’ motivation and achievement in chemistry classes [7].

1.1. Problem of Research

Generally, learners have difficulty in explaining chemical reactions and connecting between phenomena, then how to represent that [2,8,9]. Learners have many difficulties in studying microscopic and symbolic representations rather than macroscopic ones. It happens because the levels of microscopic and symbolic are invisible and abstract so it requires reasoning process [10,11,12]. Assessment is a vital component in education. The interaction between assessment, curriculum, and instruction is very important to improve the teaching and learning process [13]. The research conducted by Setiadi [14] shows that at planning stage, many teachers ignore the function of pre-test and do not perform an instrument analysis before the assessment process. Teachers also find difficulties in getting the results of assessment to know student learning progress as well as student learning difficulties [15].

Credibility of the research refers to how accurately to answers research questions or the strength of research conclusions. Indicators of success in measurement instruments are reliability and validity of measurement. Validity refers to measurement accuracy. Validity refers to how well the assessment tool is able to measure the desired end result [16,17]. Validity is used by researchers as evidence collected in the quality of assessment instruments. Research reliability refers to the how accurate a research can answer research questions and the validity of the research conclusion. The validity refers to the measurement accuracy. It relies on how good measurement tools can measure the final result [17]. It is used by the researcher as the collected evidence in the form of measurement instrument quality. Reliability shows that multiple choices measure something consistently. Reliability does not what kind of knowledge, ability, and/or skill that is measured. Therefore, the evidence of
validity becomes an important aspect before concluding that the order of multiple choices is valid [18]. The quality of multiple choices is determined by field trial process to evaluate the characteristics of each item. In fact, many teachers do not perform an instrument analysis before the assessment process. Many teachers never do pre-test and multiple choice analysis because they do not have the competency to analyze the test [19]. Teacher develops assessment test instrument only to complete content they will give to the students. The test only focuses on the numbers arrangement. It does not stimulate the way how students has to solve the problem. Consequently, the ability of students to think high, critical, creative, and problem solving is not too prominent. It only plays on numbers [20].

1.2. Research Focus

The study of the multiple representation instrument of test arrangement is important because it is useful to measure and analyze the students’ ability of multiple representations on electrolyte and nonelectrolyte materials. It helps teachers to use appropriate strategies, approaches, or learning models. This research gives many benefits to students, teachers, schools and stakeholders. It needs good and credible instrument and of course completed validation and reliability to measure students’ multiple representation ability. This research aims to validate the instrument test to measure multiple representations of senior high school students’ ability in Yogyakarta on electrolyte and nonelectrolyte materials. The research problem discussed relates to the material of electrolyte and nonelectrolyte solutions only.

2. Material and Methods

2.1. Research Design

This research is a quantitative descriptive research by using sample of senior high students with the same tenth grade for electrolyte and nonelectrolyte material. It uses non-experiment design where researcher does not give special treatment to the students.

2.2. Sample of Research

The population of the study is all senior high school tenth grade students in Yogyakarta. The sample is taken using stratified purposive sampling method from 3 senior high schools consisting of 94 students as research sample.

2.3. Instrument and Procedures

The tested multiple representation instruments consist of 40 multiple choice items covering the macroscopic, microscopic, symbolic, and mathematical aspects given after the learning process. The instrument are completed with answer sheets, manual, and key answer. The item of instrument are arranged systematically by paying attention to comprehensiveness and depth of the material.

Table 1. Grid of multiple representation instruments for electrolyte and nonelectrolyte materials

| Item Number | Aspect of Representation | Indicators of Competence Achievement | Item Indicator |
|-------------|--------------------------|--------------------------------------|----------------|
| 1 | Macroscopic Microscopic | Analyze the symptoms that happen in the tool set | Students are able to analyze microscopic representation from macroscopic images of solution and symptoms that happen on electric conductive test equipment. |
| 2 | Macroscopic Microscopic | | Students are able to analyze Macroscopic phenomena by analyzing Microscopic representation of a solution. |
| 3 | Macroscopic Microscopic | | Students are able to identify a substance by analyzing Microscopic and Macroscopic representations from the questions. |
| 4 | Macroscopic Microscopic | | Students are able to analyze symptoms of solution that should happen by observing Macroscopic and Microscopic representations of the solution. |
| 5 | Macroscopic Microscopic | Categorize the characteristics of some solutions in the environment and the solutions in the laboratory | Students are able to categorize the characteristics of a solution by analyzing the Macroscopic and Microscopic phenomena. |
| 6 | Macroscopic Microscopic | | Students are able to categorize the characteristics of a solution by analyzing the Microscopic representation. |
| 7 | Macroscopic Microscopic | Interpret images to categorize the meaning of the solution into strong electrolytes, weak electrolytes, and nonelectrolytes. | Students are able to interpret Macroscopic and Microscopic representations of a solution in order to find the cause of the solution in conducting an electric current. |
| 8 | Macroscopic Microscopic | | Students are able to interpret a Microscopic representation of a solution in order to explain the Macroscopic phenomenon. |
| 9 | Macroscopic Microscopic | | Students are able to interpret Macroscopic and Microscopic representations of some solutions in order to define the meaning of the solution for strong electrolytes and nonelectrolytes. |
| 10 | Microscopic Symbolic | Analyzes strong electrolyte solutions, weak electrolytes, and nonelectrolytes based on their electrical conductivity. | Students are able to analyze the power of different electrical conductivities by analyzing some solutions in terms of its Macroscopic and Microscopic representations. |
| 11 | Microscopic Symbolic | | Students are able to analyze different Microscopic representations of some solutions based on its electrical conductivity. |
| 12 | Microscopic Symbolic | | Students are able to analyze the characteristics of electrolyte solution compound based on Microscopic observation. |
| 13 | Microscopic Symbolic | | Students are able to analyze the characteristics of strong electrolyte solution compounds based on Microscopic observation. |
| Item Number | Aspect of Representation | Indicators of Competence Achievement | Item Indicator |
|-------------|--------------------------|---------------------------------------|---------------|
| 14          | Macroscopic Microscopic  | Analyze the mechanism of electrolyte solution in conducting electricity. | Students are able to analyze the bubbles formed from the process of an electrolyte solution in conducting electricity. |
| 15          | Macroscopic Microscopic  | Analyze the mechanism of electrolyte solution in conducting electricity. | Students are able to analyze the causes of dissolved solids in water presented in Macroscopic and Microscopic representations that can conduct electricity. |
| 16          | Macroscopic Microscopic  | Analyze the mechanism of electrolyte solution in conducting electricity. | Students are able to analyze the mechanism of electrolyte solution in conducting electricity presented in Macroscopic and Microscopic representation. |
| 17          |                         |                                       |               |
| 18          | Symbolic Microscopic     | Analyze the type of chemical bond and the electrolyte characteristics of substance. | Students are able to analyze the type of chemical bond and the electrolyte characteristics of substance presented in symbolic and Microscopic representations. |
| 19          | Macroscopic Microscopic  | Analyze the type of chemical bond and the electrolyte characteristics of substance. | Students are able to analyze the Lewis structure from the chemical bond types and the electrolyte characteristics of substance. |
| 20          | Symbolic Microscopic     | Concludes that the electrolyte solution can be ionic compound or polar covalent compound. | Students are able to analyze the causes of covalent compounds that can conduct electricity. |
| 21          | Macroscopic Microscopic  | Concludes that the electrolyte solution can be ionic compound or polar covalent compound. | Students are able to analyze the cause of phase differences in ionic compounds that affect the electrical conductivity of the electrolyte solution. |
| 22          | Macroscopic Microscopic  | Concludes that the electrolyte solution can be ionic compound or polar covalent compound. | Students are able to conclude that the electrolyte solution can be ionic compound or polar covalent compound from Macroscopic, Microscopic, and symbolic representations. |
| 23          | Macroscopic Microscopic  | Concludes that the electrolyte solution can be ionic compound or polar covalent compound. | Students are able to analyze the causes of covalent compounds that can conduct electricity. |
| 24          | Microscopic Symbolic     | Describe the ionisation of compound. | Students are able to describe the ionization of compound. |
| 25          | Microscopic Symbolic     | Describe the ionisation of compound. | Students are able to describe the ionization of compound. |
| 26          | Microscopic Symbolic     | Describe the ionisation of compound. | Students are able to analyze the type of the compound from the result of ionization. |
| 27          | Microscopic Symbolic     | Describe the ionisation of compound. | Students are able to analyze the type of the compound from the result of ionization. |
| 28          | Macroscopic Mathematics  | Calculate the volume and moles of a solution. | Students are able to calculate the volume of a substance with given measurement. |
| 29          | Symbolic Mathematics     | Calculate the volume and moles of a solution. | Students are able to calculate the mole of ionized solution easily. |
| 30          | Microscopic Mathematics  | Calculate the volume and moles of a solution. | Students are able to calculate the degree of ionization easily. |
| 31          | Macroscopic Microscopic  | Predict the electrical conductivity of a solution. | Students are able to analyze concentration effect on Macroscopic and Microscopic representations in order to predict the electrical conductivity of a solution. |
| 32          | Macroscopic Microscopic  | Predict the electrical conductivity of a solution. | Students are able to analyze the effect of concentration on Macroscopic and Microscopic representation for predicting the order of power conductivity of a solution. |
| 33          | Macroscopic Microscopic  | Predict the electrical conductivity of a solution. | Students are able to analyze the effect of ions number presented in Macroscopic and Microscopic representations in order to predict the electrical conductivity of a solution. |
| 34          | Macroscopic Microscopic  | Predict the electrical conductivity of a solution. | Students are able to analyze the effect of ions number presented in Macroscopic and Microscopic representations for predicting the order of power from electrical conductivity of a solution. |
| 35          | Macroscopic Microscopic  | Critize the image of solution characteristics and electrical conductivity. | Students are able to criticize the image of solution characteristics and the electrical conductivity from Macroscopic and Microscopic representations. |
| 36          | Macroscopic Microscopic  | Conclude the function of electrolyte solution in human body. | Students are able to determine the electrolyte substances in human body. |
| 37          | Microscopic Macroscopic  | Conclude the function of electrolyte solution in human body. | Students are able to find the function of electrolyte solution in human body. |
| 38          | Macroscopic Macroscopic  | Conclude the function of electrolyte solution in human body. | Students are able to find the function of electrolyte solution in human body. |
| 39          | Macroscopic Symbolic     | Decide how to deal with electrolyte deficiency in human body. | Students are able to make decisions on how to deal with electrolyte deficiency in human body. |
| 40          | Macroscopic Symbolic     | Decide how to deal with electrolyte deficiency in human body. | Students are able to make decisions on how to deal with electrolyte deficiency in human body. |

### 2.4. Data Analysis

The aims of this research is to validate the instrument test so the validation was done in two ways namely theoretical validation and empirical validation. For theoretical validation performed by expert judgement and education practitioners. The instruments that have been compiled are validated by validators using V-Aiken. For empirical validation done by testing the instrument to 94 students from 3 schools in Yogyakarta city. Empirical validation of the instruments is analyzed using Rasch Model to determine the validity, reliability, fit items, problem level, and function of information.
3. Results

The application of Rasch model in this research in determining instrument validity and reliability is important to define valid and reliable item construct and to give a clear definition about measurable construction in consistent with theory. Interestingly, this model can be used effectively for consistently measurable items and valid response pattern [6].

3.1. Theoretical Validation

The instrument of the test consists of 40 items of multiple choice with pretest and key answer. The judgment of experts and education practitioners performs as theoretical validation for the questions that have been arranged in the aspect of content, construction, and language. Then the result of validation is analyzed using V-Aiken. Table 2 shows the results of Aiken index analysis.

The result of expert judgment and education practitioner judgement (4 rater in total) and 5 rating scale category in Table 2 shows that the average index of Aiken is 0.956 for the substance aspect, 0.967 for the construction aspect and 0.959 for the language aspect. Item with index more than 0.88 is acceptable. The result of the analysis indicates that total Aiken index from 40 item has an average more than 0.88. It can be concluded that the multiple representation instrument fulfills the theoretical validity and can be used for empirical validity testing [21].

3.2. Empirical Validation

Empirical validation is performed to construct validation and decision of sample measure by using SPSS 16 Program. The subject of empirical validation in this research is 94 tenth grade students of 3 senior high schools in Yogyakarta. The first phase is to do sample adequacy test. The measure (size) or number of samples is sufficient and eligible for analysis when the analysis of KMO-MSA test is more than 0.05 (> 0.50) and Bartlett test significance is less than 0.01 (<0.01). Table 3 shows the result of sample adequacy test.

Table 2. Aiken index of content, construction, and language aspects

| Item Number | Aiken Index of Content Aspect | Aiken Index of Construction Aspect | Aiken Index of Language Aspect | Item Number | Aiken Index of Content Aspect | Aiken Index of Construction Aspect | Aiken Index of Language Aspect |
|-------------|-------------------------------|-----------------------------------|-------------------------------|-------------|-------------------------------|-----------------------------------|-------------------------------|
| 1           | 1.000                         | 0.938                             | 0.813                         | 21          | 1.000                         | 1.000                             | 1.000                         |
| 2           | 1.000                         | 1.000                             | 0.938                         | 22          | 1.000                         | 0.938                             | 0.875                         |
| 3           | 1.000                         | 1.000                             | 1.000                         | 23          | 1.000                         | 1.000                             | 0.938                         |
| 4           | 1.000                         | 0.813                             | 0.875                         | 24          | 0.875                         | 0.938                             | 1.000                         |
| 5           | 1.000                         | 0.875                             | 1.000                         | 25          | 0.875                         | 1.000                             | 1.000                         |
| 6           | 1.000                         | 0.875                             | 0.938                         | 26          | 0.813                         | 0.938                             | 1.000                         |
| 7           | 0.938                         | 0.938                             | 1.000                         | 27          | 0.938                         | 1.000                             | 1.000                         |
| 8           | 1.000                         | 1.000                             | 0.875                         | 28          | 1.000                         | 0.963                             | 1.000                         |
| 9           | 1.000                         | 1.000                             | 0.938                         | 29          | 0.938                         | 1.000                             | 1.000                         |
| 10          | 0.938                         | 0.938                             | 1.000                         | 30          | 0.875                         | 1.000                             | 0.938                         |
| 11          | 1.000                         | 1.000                             | 0.875                         | 31          | 1.000                         | 0.938                             | 0.813                         |
| 12          | 0.875                         | 0.875                             | 0.938                         | 32          | 1.000                         | 1.000                             | 1.000                         |
| 13          | 0.938                         | 0.938                             | 0.938                         | 33          | 1.000                         | 1.000                             | 0.938                         |
| 14          | 0.938                         | 1.000                             | 0.988                         | 34          | 1.000                         | 1.000                             | 0.938                         |
| 15          | 1.000                         | 1.000                             | 1.000                         | 35          | 0.938                         | 1.000                             | 1.000                         |
| 16          | 1.000                         | 1.000                             | 1.000                         | 36          | 0.938                         | 1.000                             | 0.938                         |
| 17          | 1.000                         | 1.000                             | 0.875                         | 37          | 0.938                         | 1.000                             | 1.000                         |
| 18          | 1.000                         | 0.938                             | 0.938                         | 38          | 0.938                         | 1.000                             | 1.000                         |
| 19          | 0.875                         | 1.000                             | 0.938                         | 39          | 0.938                         | 1.000                             | 1.000                         |
| 20          | 0.875                         | 1.000                             | 1.000                         | 40          | 0.875                         | 1.000                             | 0.938                         |
| Average     | 0.956                         | 0.967                             | 0.953                         |             |                               |                                   |                               |

Table 3. The result of KMO-MSA and Bartlett tests

| Test Analysis        | Test Result | Criteria | Conclusion                      |
|----------------------|-------------|----------|---------------------------------|
| KMO-MSA              | 0.694       | >0.50    | Sample measure (size) is eligible for further analysis. |
| Bartlett Significance Test | 0.00 | <0.01    |                                    |
3.2.1. Unidimension

Unidimension is defined as the presence of a latent underlying data trait [22]. Unidimension of the instrument is an important measure to evaluate whether the instrument can measure multiple representational abilities of students from dimension items of Winstep Program.

According to Figure 1, the result of raw variance measurement data is 31.7%. It indicates that the test instruments is only able to measure 31.7% of the desired capability with 68.3% of data that is unexplainable. The minimum unidimensionality requirement is 20% [23] and it can be concluded that the instrument is only able to measure one representation of multiple capabilities [24].

3.2.2. Item Fit

Item fit explains whether the items work normally to do measurement or not. The item fit analysis is done using Winstep program. Table 4 shows the criteria of conformity level of the items according to Sumintono & Widhiarso [25].

The test instrument has to meet one criteria to be considered fit. When it does not meet one criteria, e.g. it does not meet the criteria of MNSQ outfit then it can be rematched with the criteria of ZSTD or Pt. Mean Corr outfit [25]. Table 5 shows the result of item fit analysis of this research.

![Table 4. The criteria of fit item](image)

| Item Fit                  | Criteria                      |
|--------------------------|-------------------------------|
| Outfit Mean Square (MNSQ)| 0.5 < MNSQ < 1.5              |
| Outfit Z-Standard (ZSTD) | -2 < ZSTD < +2                |
| Point Measure Correlation (Pt Mean Corr) | 0.4 < Pt. Measure Corr < 0.85 |

![Table 5. The result of item fit analysis](image)

| Item Number | MNSQ Outfit | ZSTD Outfit | Pt. Mean Corr Outfit | Result  | Item Number | MNSQ Outfit | ZSTD Outfit | Pt. Mean Corr Outfit | Result  |
|-------------|-------------|-------------|----------------------|---------|-------------|-------------|-------------|----------------------|---------|
| 1           | 0.54        | -0.9        | 0.43                 | fit     | 21          | 0.90        | -0.7        | 0.48                 | fit     |
| 2           | 0.85        | -0.4        | 0.42                 | fit     | 22          | 0.86        | -0.5        | 0.46                 | fit     |
| 3           | 0.84        | -0.4        | 0.40                 | fit     | 23          | 0.83        | -1.2        | 0.53                 | fit     |
| 4           | 0.87        | -0.2        | 0.36                 | fit     | 24          | 1.17        | 0.32        | 0.43                 | fit     |
| 5           | 1.50        | 1.4         | 0.26                 | fit     | 25          | 0.79        | -0.7        | 0.49                 | fit     |
| 6           | 0.76        | -0.3        | 0.40                 | fit     | 26          | 0.94        | -0.4        | 0.43                 | fit     |
| 7           | 1.00        | 0.2         | 0.36                 | fit     | 27          | 1.21        | 1.3         | 0.34                 | fit     |
| 8           | 1.03        | 0.3         | 0.31                 | fit     | 28          | 0.60        | -1.1        | 0.51                 | fit     |
| 9           | 0.23        | -1.0        | 0.43                 | Does not fit | 29          | 0.99        | 0.0         | 0.41                 | fit     |
| 10          | 0.98        | 0.1         | 0.34                 | fit     | 30          | 1.82        | 4.9         | 0.16                 | Does not fit |
| 11          | 0.65        | -0.4        | 0.36                 | fit     | 31          | 0.75        | -1.5        | 0.56                 | fit     |
| 12          | 1.19        | 1.3         | 0.29                 | fit     | 32          | 0.68        | -1.9        | 0.61                 | fit     |
| 13          | 1.02        | 0.2         | 0.37                 | fit     | 33          | 0.92        | -0.3        | 1.50                 | fit     |
| 14          | 0.82        | -0.5        | 0.43                 | fit     | 34          | 1.23        | 1.1         | 0.24                 | fit     |
| 15          | 1.12        | 0.8         | 0.38                 | fit     | 35          | 1.03        | 0.2         | 0.42                 | fit     |
| 16          | 1.55        | 2.3         | 0.18                 | Does not fit | 36          | 1.18        | 1.0         | 0.29                 | fit     |
| 17          | 0.99        | 0.0         | 0.45                 | fit     | 37          | 1.17        | 0.8         | 0.35                 | fit     |
| 18          | 1.02        | 0.2         | 0.41                 | fit     | 38          | 0.96        | -0.1        | 0.42                 | fit     |
| 19          | 1.04        | 0.4         | 0.36                 | fit     | 39          | 0.86        | -0.9        | 0.50                 | fit     |
| 20          | 0.77        | -1.6        | 0.54                 | fit     | 40          | 0.99        | 0.0         | 0.37                 | fit     |
Table 5 shows the results of item fit analysis to measure the students' ability of multiple representations. According to information of Table 5, there are 3 unfit items from total 40 items after analysis process using Rasch model PCM 1-PL. The 3 unfit items are 9, 16 and 30. The unfit items are removed for the misconception of students on the items, leaving 37 items to measure the ability of multiple representations of students.

3.2.3. Test Reliability

Reliability refers to whether test instrument gives the same result every time it is used in the same setting with the same subject type. Basically, reliability means consistent or reliable results. Reliability is a part of the validity assessment [17]. The analysis reliability test considers the value of Alpha Cronbach that is obtained from Winstep program analysis; summary statistics. According to Bhatnagar et al. [26], it uses the criteria to see reliability based on Alpha Cronbach value that can be seen in Table 6.

Table 6. The criteria of reliability

| Alpha Cronbach Value | Category       |
|----------------------|---------------|
| < 0.5                | Unacceptable  |
| 0.5 < α < 0.6        | Poor          |
| 0.6 < α < 0.7        | Acceptable    |
| 0.7 < α < 0.9        | Good          |
| α > 0.9              | Excellent     |

The Alpha Cronbach value of this study is 0.86. It shows that the category of reliability in the test instrument to measure the multiple representation capabilities of students is categorized as Good [26]. The consistency of items can be used for other samples that have identical (or almost identical) characteristics [27].

3.2.4. Difficulty Level of Item (Question)

According to classical test theory, item difficulty is the percentage of students who answer an item correctly. The greater the percentage of students who work on the item correctly, the easier the item (question) is being worked on. If the test items are very difficult, then the majority of exam scores will be very low. If the test items are done very easily, then the test score will be very satisfying [28]. Item measure information is analyzed using Winstep program to find the difficulty level of items. Table 7 shows the criteria of problem level according to Adedoyin & Mokobi [7].

Table 7. The criteria of item difficulty

| Criteria | Category |
|----------|----------|
| < -1     | Easy     |
| -1 < b < +1 | Moderate |
| > +1      | Difficult |

Table 8 shows the results of difficulty item in this study.

| Item Number | Difficulty Level | Category | Item Number | Difficulty Level | Category |
|-------------|------------------|----------|-------------|------------------|----------|
| 1           | -1.15            | Easy     | 21          | 0.59             | Moderate |
| 2           | -0.91            | Moderate | 22          | -0.48            | Moderate |
| 3           | -0.82            | Moderate | 23          | 1.17             | Difficult|
| 4           | -1.25            | Easy     | 24          | 0.42             | Moderate |
| 5           | -1.01            | Easy     | 25          | -0.64            | Moderate |
| 6           | -1.91            | Easy     | 26          | 0.75             | Moderate |
| 7           | -1.75            | Easy     | 27          | 0.31             | Moderate |
| 8           | -2.32            | Easy     | 28          | -1.22            | Easy     |
| 9           | -2.92            | Easy     | 29          | 0.59             | Moderate |
| 10          | -1.46            | Easy     | 30          | 1.07             | Difficult|
| 11          | -2.10            | Easy     | 31          | 0.25             | Moderate |
| 12          | 0.70             | Moderate | 32          | 0.13             | Moderate |
| 13          | 0.25             | Moderate | 33          | -0.13            | Moderate |
| 14          | -0.82            | Moderate | 34          | 1.97             | Difficult|
| 15          | 0.36             | Moderate | 35          | 0.59             | Moderate |
| 16          | 2.03             | Difficult| 36          | 1.80             | Difficult|
| 17          | 1.01             | Difficult| 37          | 2.27             | Difficult|
| 18          | 0.19             | Moderate | 38          | -0.26            | Moderate |
| 19          | 1.22             | Difficult| 39          | 0.53             | Moderate |
| 20          | 1.53             | Difficult| 40          | 1.97             | Difficult|

Table 8 provides information about the difficulty level of items from the instrument. The analysis results of the difficulty level of items in the instrument that measures students' ability of multiple representations shows that there are 10 easy items, 20 medium items, and 10 difficult items.

3.2.5. Test Information Function

Test information function is performed to obtain information of the measurements. The information function is analyzed using the Winstep program that can be performed from the menu of graph-information function. Figure 4 shows the curve of test information function from this study.
According to the curve in Figure 4, it can be concluded that the result of 40 questions given to 94 students represents a very low ability levels, and the information from the measurement is also quite low. Similarly, at the very high ability level the information from the measurement is also quite low. At moderate ability level, the information from the measurement is very high. The conclusion shows that well arranged items determine the students' ability moderate level only [24].

4. Discussion

The result of this study is theoretical validity analyzed by expert judgement and education practitioner judgement (4 rater in total) and 5 rating scale category shows that the average index of Aiken is 0.991 for the substance and language aspects, and 0.993 for the construction aspect. The highest Aiken index is 1.000 and the lowest is 0.963. Item with index more than 0.88 is acceptable. The result of the analysis indicates that total Aiken index from 40 item questions is more than 0.88. It can be concluded that the multiple representation instrument fullfills the theoretical validity and can be used for empirical validity testing [21].

The next step is empirical validation. The first phase is to do sample adequacy test. The result of KMO-MSA test analysis is 0.694 and Bartlett Significance Test is 0.00. It concludes that the sample size of test is eligible for further analysis. The second phase is unidimensional analysis. Unidimension is defined as the presence of a latent underlying data trait. The result shows that the instrument is only able to measure one representation of multiple capabilities [24]. The third phase is checking item fit. The results is 3 unfit items are number 9, 16 and 30 from total 40 items after analysis process using Rasch model PCM 1-PL. The unfit items are removed for the misconception of students on the items, leaving 37 items to measure the ability of multiple representations of students. The fourth phase is measure item and person reliability. The Alpha Cronbach value of this study is 0.86. It shows that the category of reliability in the test instrument to measure the multiple representation capabilities of students is categorized as Good. Item and person reliability showed that It can be concluded that both consistency of respondents answers and quality of items in special instruments are good. The fifth phase is analysis difficulty level of items. The analysis results of the difficulty level of items in the instrument that measures students’ ability of multiple representations shows that there are 10 easy items, 20 medium items, and 10 difficult items. The most difficult item is number 37 and the most easy item is number 9. At moderate ability level, the information from the measurement is very high. The conclusion shows that well arranged items determine the students' ability level only [24].

5. Conclusions

The instrument of assessment must be reliable and valid to get credible measurement results. Thus, the reliability and validity of each assessment instrument that is used to measure the results of the study should be analyzed. Validity refers to how accurate a study answers research questions or the strength of research conclusions [17]. The instrument of test has fulfilled the theoretical validity performed by the expert judgement and educational practitioners who have been analyzed using the V-Aiken index. The result of the analysis shows that the test instruments have fulfilled theoretical validity and can be used for further analysis. The arranged test instruments
have fulfilled the empirical validity with empirical evidence of 37 fit items from a total of 40 items that have been analyzed using PCM 1-PL. The reliability of test instruments categorized as good [26]. Based on the value of person reliability and item reliability, it can be concluded that the consistency of answers from respondents and quality items in a special instrument are good. It indicates that the arranged instruments can be categorized as good questions. The test instrument can be used as a tool to diagnose or to measure students’ ability of multiple representations on electrolyte and nonelectrolyte solutions. The arranged items determine the students' ability moderate level only [24].

Suggestion for next research is that chemistry teachers can apply the instruments that have been prepared to determine students’ ability of multiple representations to help teachers to decide the best strategy, approach, or appropriate learning models then students understand multiple representations in learning electrolyte and nonelectrolyte solutions. This instrument can be used as an example to prepare test instruments for other chemical contents.

Acknowledgements

The authors are thankful to all teachers and students at state senior high school in Yogyakarta, many thanks for their supported on this study.

References

[1] Talanquer, V., “Macro, submicro, and symbolic: The many faces of the chemistry “triplet”,” International Journal of Science Education, 33 (2), 179-195. 2011.
[2] Gabel, D, “Improving teaching and learning through chemistry education research: A look to the future,” Journal of Chemical Education, 76 (4), 548-554. 1999.
[3] Hernandez, G.E., Criswell, B.A., Kirk, N.J., Sauer, D.G., & Rushton, G.T, “Pushing for particulate level models of adiabatic and isothermal processes in upper-level chemistry courses: a qualitative study,” The Royal Society of Chemistry: Chemistry Education Research and Practice, 15, 354-365. 2014.
[4] Wu, H.K., & Shah, P, “Exploring visuospatial thinking in chemistry learning,” Science Education, 88 (3), 465-492. 2003.
[5] Johnstone, A.H, “The development of chemistry teaching a changing response to changing demand,” The Forum Symposium on Revolution and Evolution in Chemistry Education, 70 (9), 701-705. 1993.
[6] Milenkovic, D.D., Segedinac, M.D., & Hrin, T.N, “Increasing high school students’ chemistry performance and reducing cognitive load through an instructional strategy based on the interaction of multiple levels of knowledge representation,” Journal of Chemical Education, A-H. 2014.
[7] Adeoyin, O.O., & Mokobi, T, “Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items,” International Journal of Asian Social Science, 3 (4), 992-1011. 2013.
[8] Ainsworth, S, “The functions of multiple representations,” Computers & Education, 33 (2-3), 151-152. 1999.
[9] Gilbert, J.K, & Treagust, D, Models and modeling in science education: Multiple representations in chemical education, Springer, Perth, 2009.
[10] Demircioglu, G, Demircioglu, H, & Yadigaroglu, M, “An investigation of chemistry student teachers’ understanding of chemical equilibrium,” International Journal on New Trends in Education and Their Implications, 4 (2), 185-192. 2013.
[11] Chandrasegaran, A. L., Treagust, D.F., & Mocerino, M, “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. 2007.
[12] Griffiths, A.K., & Preston, K.R, “Grade-12 students’ misconceptions relating to fundamental characteristics of atoms and molecules,” Journal of Research in Science Teaching, 29 (6), 611-628. 1992.
[13] Ghazali, N.H.M, “A reliability and validity of an instrument to evaluate the school-based assessment system: a pilot study,” Journal of Evaluation and Research in Education, 5 (2), 148-157. 2016.
[14] Setiadi, H, “Pelaksanaan penilaian pada kurikulum 2013,” Jurnal Penelitian dan Evaluasi Pendidikan, 20 (2), 166-178. 2016.
[15] MaisaryOh, Zulkarnain W, Setyowati A.J., & Mahanal, S, “Masalah guru dalam implementasi kurikulum 2013 dan kerangka model supervisi pengajaran,” Manajemen Pendidikan, 24 (3), 213-220. 2014.
[16] Kimberlin C.L., & Winterstein, A.G, “Validity and reliability of measurement instruments used in research,” American Society of Health System Pharmacists, 65, 2276-2284. 2008.
[17] Sullivan, G.M, “A primer on the validity of assessment instruments,” Journal of Graduate Medical Education, 119-120. 2011.
[18] Peeters, M.J., Beltuyukova, S.A., & Martin, B.A, “Educational testing and validity of conclusions in the scholarship of teaching and learning,” American Journal of Pharmaceutical Education, 77 (9), 1-9. 2013.
[19] Sanova, A., Bakar, A., & Afrida, “Standarisisasi instrumen penilaian hasil belajar dengan program anates v4 bagi-guru smnp 17 kota Jambi,” Jurnal Pengabdian Masyarakat, 2 (1), 1-10. 2017.
[20] Baharudin, “Menganalisis instrumen penilaian pembelajaran matematika pada materi segi empat sekolah menengah pertama negeri i Dompulu,” Jurnal Kependidikan, 15 (1), 1-10. 2013.
[21] Aiken, L.R, “Content validity and reliability of single items or questionnaires,” Educational and Psychological Measurement, 40 (4), 955-959. 1980.
[22] Hattie, J, “Methodology review: assessing unidimensionality of test and items,” Applied Psychological Measurement, 9 (2), 139-164. 1985.
[23] Reckase, M.D, “Unifactor latent trait models applied to multifactor test: results and implications,” Journal of Educational Statistics, 4 (3), 207-230. 1979.
[24] Sumintono, B. & Widiario, W, Aplikasi pemodelan rasch: pada assessment pendidikan. Trim Komunikata Publishing House, Cimahi, 2015, 78-122.
[25] Sumintono, B. & Widiario, W, Aplikasi model rasch untuk penelitian ilmu-ilmu sosial (Rev. ed). Trim Komunikata Publishing House, Cimahi, 2014, 072.
[26] Bhatnagar, R., Kim, J., & Many J.E, “Candidate surveys on program evaluation: examining instrument reliability, validity and program effectiveness,” American Journal of Educational Research, 2 (8), 683-690. 2014.
[27] Shah, R.L.Z.R.M., Samad, M.H.A., Shah, R.N.F.A.R.M., Adenan, N.H, “Validity and reliability of graphing calculator skills test items for circles topic (CGCST) using Rasch measurement model analysis: a pilot study,” International Journal of Education and Research, 5 (8), 189-200. 2017.
[28] Afolabi, E.R.I., Owolabi, C.O., & Iwintolu, R.O, “Validation of osborn’s scale for measuring the relative difficulty of secondary school subjects,” International Journal of Education and Research, 4 (2), 139-148. 2016.