Jurnal Pendidikan Sains Universitas Muhammadiyah Semarang.
October Vol. 10(2) pp 32-43 ISSN:2339-0786
DOI: https://doi.org/10.26714/jps.10.2.2022.32-43

JPS
JOURNAL
PENDIDIKAN SAINS
JURNAL
PENDIDIKAN SAINS
p-ISSN : 2339-0786 o-ISSN : 2502-1443
UNIVERSITAS MUHAMMADIYAH SEMARANG

Preliminary Analysis: Development of Higher Order Thinking Skills Problem on Stoichiometry Topic Using the Rasch Model

Abdullah [a,1,*], Rasmiwetti [b,2], Ardiansyah [c,3]

[a,b] Department of Chemistry Education, University of Riau, Pekanbaru, 28293, Indonesia
[c] Department of Chemistry Education, State Islamic University of Sultan Syarif Kasim Riau, Pekanbaru, 28293, Indonesia
[1] abdoel71@gmail.com *; [2] rasmiwetti.19@gmail.com ; [3] ardiansyahm.pd@uin-suska.ac.id
* corresponding author

Article history
Submission : 2022-01-26
Revised : 2022-10-21
Accepted : 2022-11-02

Keyword
Rasch Model
higher-order thinking skills
item analysis
stoichiometry

Abstract

This study is a part of the development study of stoichiometric HOTS problems. This study aimed to develop the stoichiometry Higher Order Thinking Skills (HOTS) test instrument. The stages of the study started with analyzing the material on the stoichiometry topic. The test consisted of 10 multiple-choice questions with 5 answer options. The instrument test was assessed by two chemistry education experts before being administered to 26 high school students in Pekanbaru city. The test results were analyzed with the Rasch Measurement Model (RMM) using the Win Step software. RMM was useful to investigate item validity, item and person reliability, and item-person map. All of these criteria were useful in improving the quality of the instrument based on Item Response Theory. The results showed that almost all items were valid, except for the Q16 items. Thus, these nine items could be used as higher-order thinking test instruments for a larger sample.

1. INTRODUCTION

The 2013 curriculum emphasizes the achievement of students' higher-order thinking skills (HOTS). HOTS are the ability of students to connect the knowledge gained with the reality of their lives (Kusuma et al., 2017). Students are expected to be able to use their mindsets to solve complex real-world problems. HOTS are activated when students encounter unfamiliar problems, uncertainties, questions, or dilemmas (Saido et al., 2015). Successful application of HOTS in the classroom will make students' intellectual continue to grow. This will make HOTS an extremely required skill for every individual in any educational setting (Tan & Halili, 2015). Therefore, the development and assessment of students’ HOTS are pivotal to be achieved to prepare students for the future.

HOTS can be developed by applying appropriate learning models and strategies. Sarif et al. (2019) found that project-based learning through recycling waste cooking oil can improve students' HOTS effectively. HOTS can also be enhanced by applying problem-based learning (Jailani et al., 2017) and discovery learning models (Cola & Nuswowati, 2020). In addition, HOTS can be improved by implementing an inductive reasoning strategy (Misrom et al., 2020) and laboratory activity (Lisdiani et al., 2019). Although many models and strategies can be used to improve students' HOTS, the Indonesian Educational Monitoring Network (JPPI) states that its application by teachers in the classroom is not evenly distributed and is not optimal. Ramdiah et al., 2019 explained that high school biology teachers in Banjarmasin have prepared to learn well, but this is not carried out consistently, and HOTS aspects in teaching tend to be ignored. Equally important in developing students’ HOTS is the construction of appropriate HOTS assessment instruments.

The competence of teachers in developing HOTS assessment instruments is still considered low (Dahlan et al., 2020). The main factor causing the low competence in the construction of HOTS problems is the lack of teachers’ knowledge and experience (Tyas et al., 2019). The evaluation developed by the teacher has not yet assessed students to think critically and creatively. As a result, the learning objectives embedded in K-13 have not been achieved. The 2013 curriculum requires that assessments conducted by teachers in schools can measure the HOTS of students (Kurniawan & Noviana, 2017). Therefore, instruments are needed to assist teachers in preparing evaluations appropriately.

Chemistry is one of the subjects taught to high school students. A topic taught in high school chemistry is stoichiometry. Stoichiometry is the basic material for chemical calculations and is a prerequisite for several other materials, such as solution chemistry, chemical equilibrium, kinetics, acid-base, colligative solutions, and others. The mastery of stoichiometry material is highly emphasized in the mastery of chemical counts. If the mastery is good, then the learning achievement of the stoichiometry topic can be improved (Abdullah et al., 2019). To see the extent to which students’ mastery of the stoichiometry concept requires a valid and reliable evaluation instrument. Many tools can be used to help develop a good assessment instrument, one of which is the Rasch model.

The Rasch model is a social instrument analysis model with numerous advantages, such as providing a linear measure with the same intervals, the correct estimation process, finding incorrect items (misfits) or uncommon (outliers), overcoming missing data, and producing replicable measurements (independent of the parameters studied) (Bond, 2010). The application of the Rasch model can produce a reliable and valid instrument (Razali et al, 2016). Application of the Rasch model can plot ability scores, and item difficulty scores, and presents the correlation between the probability of accomplishing a task (Cavanagh & Waugh, 2011). The use of the Rasch model in item analysis is expected to improve the quality of the questions (Karlin & Karlin, 2018). Mohamad et al., 2015 have successfully measured the validity of the instrument with the Rasch model and stated that the Rasch Model is suitable to be applied in instrument validation.

Susongko (2016) has studied the validation of the instrument test with the Rasch model, which revealed that the instrument is proven valid using the Rasch Measurement Model (RMM). The study of item analysis for adapted motivation scales in second language learning using the Rasch Model has been conducted by Ng et al. (2018). They concluded that the instrument is valid and reliable to evaluate student motivation. In this present study, the development and analysis of stoichiometry problem items using the Rasch Model have been carried out. The instrument validation is analyzed by referring to the value of point measure correlation (PMC), infit and outfit of standardized *z* value (ZSTD), and mean square (MNSQ) to specify the validity of items, and Cronbach's alpha coefficient to determine the reliability of items and person.

The purpose of this study is to develop the HOTS problem on stoichiometry topic using Rasch Model. This study discusses the initial development of the HOTS problem on the stoichiometry topic and the analysis of the characteristics of the problem based on the Rasch model using the MiniStep software.

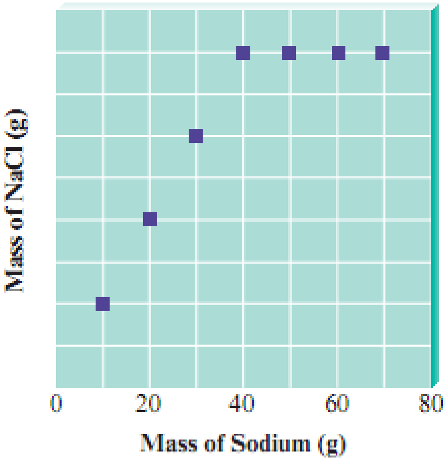
2. METHOD

The sample of this pilot study was 26 tenth graders high school students in Pekanbaru city. The students answered the Multiple Choice HOTS Problem (MCHP) on stoichiometry scope, and the results were analyzed with the Rash model using MiniStep software.

2.1. Instrument

The instruments of MCHP on the stoichiometry topic were arranged based on essential concepts on the stoichiometry topic. It consisted of ten multiple choice stoichiometric tests with a different level HOTS based on the revised Bloom taxonomy as shown in Table 1. The instrument was subsequently assessed and revised by two experts. The instrument that was developed successfully was tested on 26 high school students. The results of the trial were then analyzed with the Rasch model using MiniStep software. The MiniStep is a freeware that can be downloaded at https://www.winsteps.com/ministep.htm.

Table 1. Some developed MCHP problem on stoichiometry topic

Content Problem	Answer
Students were provided with experimental data about the chemical reaction. Students could assess the phenomena occurring concerning the basic laws of chemistry. The question was: *You have 7 closed containers, each containing the same amount of Cl_2 gas. You add 10 grams of sodium in container I, 20 grams of sodium in container II, 30 grams of sodium in container III, and so on until container VII is added 70 grams of sodium. Sodium and chlorine react to produce sodium chloride based on the following graph:*  ***Figure 1.** Graph of the mass of NaCl formed* *The most appropriate conclusion based on experimental data is…*	**D**

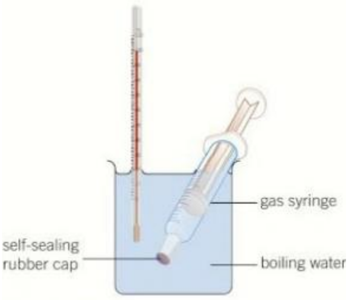
(a) The amount of NaCl formed in container II is 40 grams *(b) In containers I, II, and III there is no remaining reagent* *(c) The amount of NaCl formed in container 5 is about 90 grams* *(d) Only in the IV container there is no reagent left* *(e) There is one reagent left in all the containers*	
Students were provided molecular formulas and the relative atomic mass of a compound. Students could determine the relative molecular mass value of a compound correctly if the atomic mass value of 12C was given another value. The question was: *Epsom salt, $MgSO_4.7H_2O$, is commonly used as a detoxifying, relaxing, and smoothing agent for the skin. This salt is usually used as a mixture of water in a bath. If the relative atomic mass of ^{12}C is not 12 but 100, then the relative molecular mass of Epsom salt is….* *(a) 246* *(b) 2050* *(c) 2460* *(d) 1230* *(e) 2952*	**B**
Students were provided methods on how to determine the relative molecular simply. Students could determine the relative molecular mass of a compound correctly. The question was: *The relative molecular mass of a volatile liquid can be determined by heating the liquid in water at a temperature of 100^0C and measuring the volume of gas collected in the injection (as in Figure 2). A total of 0.2245 grams of the volatile liquid sample is heated to evaporate. If the evaporation results can fill 81 mL at 1 atm and 100^oC, then the relative molecular mass of the liquid is…*  ***Figure 2.** The molecular mass determination apparatus* *(a) 56* *(b) 65* *(c) 76* *(d) 85* *(e) 96*	**D**
Students were provided with the composition of Ringer's solution. The solution is widely used in treating wounds. Students could calculate the exact molarity of one of the compositions of Ringer's solution. The question was: *Ringer's solution is used in the treatment of wounds and burns. This solution was prepared by dissolving 4.3 grams of NaCl, 0.15 grams of KCl, and 0.165 grams of $CaCl_2$ in distilled water and diluting them to a volume of 500 mL. The molarity of the Cl^- ions in this solution is...*	**D**

Figure 3. *Ringer's solution*

(a) 0.147 M
(b) 0.00298 M
(c) 0.0042 M
(d) 0.157 M
(e) 0.23 M

2.2. Data Analysis

The preliminary analysis was conducted using MiniStep software referring to Item Response Theory (IRT) to determine the validity and reliability of MCHP on stoichiometry scope. The parameters used to test the validity and reliability of items were PMC, ZSTD, MNSQ, and the item separation index. PMC measures the index of item discrimination. Infit and MNSQ, and ZSTD outfits functions were used to determine the correlation between item difficulty and individual abilities. The same criteria were used to interpret the index of item separation and item reliability (Linacre, 2019). The research flowchart is displayed in Figure 4.

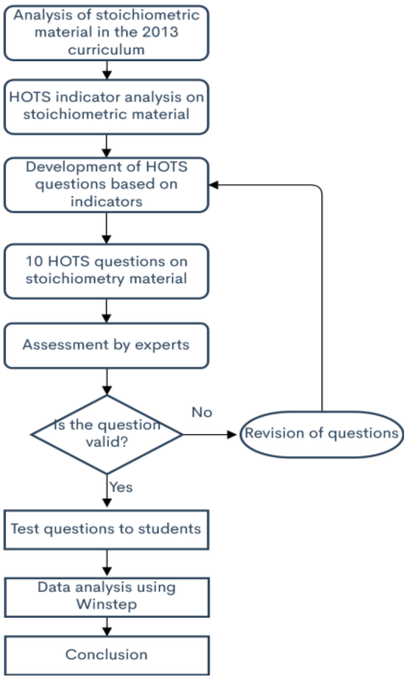


Figure 4. Research Flowchart

3. RESULTS AND DISCUSSION

The MCHP on stoichiometry scope had been declared valid by expert validators. It had been tested on 26 high school students. Validity, reliability, and other parameters were performed based on the Rasch Model using MiniStep software. The parameters to be analyzed using the MiniStep program were the item of validity and reliability, item strata, item difficulty level, item-person maps, and PMC.

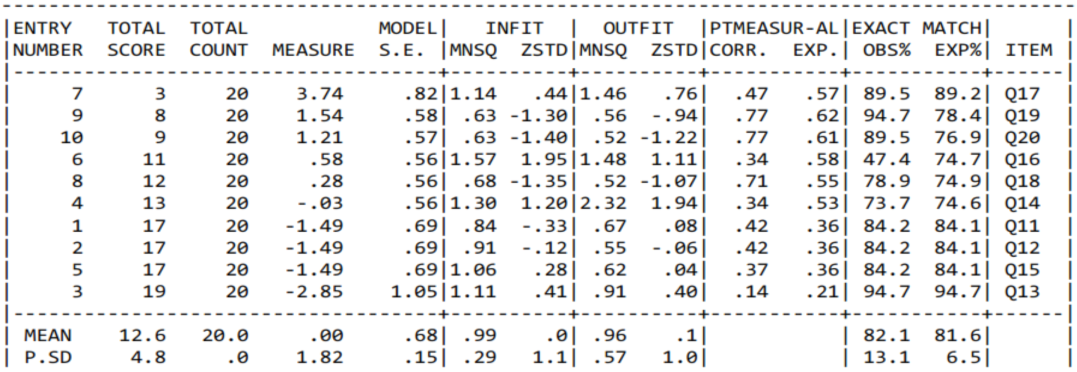
3.1. Item Validity

Validity fundamentally denotes "measure what is intended to be measured" (Field, 2009). Validity is related to the meaning of conclusions obtained from test scores (Wright & Stone, 1999). The level of item appropriateness is used to explain whether the item is functioning normally to take measurements or not. The consistency of response patterns for items and people is related to fit validity. This validity comes from the compatibility of the observed person-item responses (Wright & Stone, 1999). An item is 'fit' if the pattern of response to it is steady with the way these students have replied to the other items. If the problem found is not fit, it indicates that there is a misconception in students about the items.

Table 2 brings out the MiniStep® information, including fit statistics for all items. The items are displayed in the far-left column ('Entry Number'), followed by the number of the correct answer and the number of opportunities for each item. The 'MEASURE' column indicates the Rasch size for the item and is reported as log odd units (*logit*). It shows whether the items support one construct and the reach to which one person's response is reliable (Cordier et al., 2018). Values for MNSQ, infit and outfit of ZSTD, and PMC were observed and estimated for each item in logit.

Based on Table 2, none of the items showed a negative PMC. This means that all existing items measure the construct in the same orientation as desired. According to (Linacre, 2019), items are suitable if the MNSQ value range is 0.5-1.5. Almost all items meet these criteria, except for item Q16. It means that the construct of almost of items is precise and effective for evaluating self-regulation indicators (Habibi et al., 2019). Specifically, item Q16, this item is not appropriate and should be explored in more detail.

Table 2. Item statistic output of MCHP



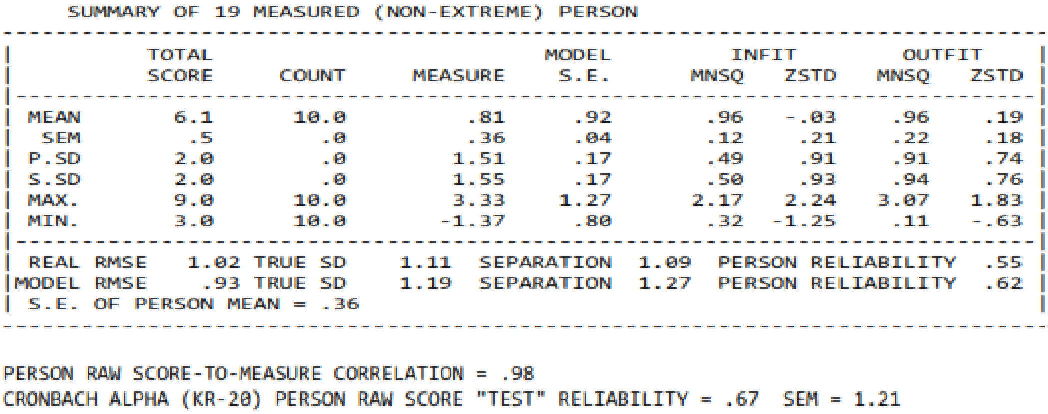
```
---------------------------------------------------------------------------------------------
|ENTRY    TOTAL   TOTAL             MODEL|    INFIT    |   OUTFIT    |PTMEASUR-AL|EXACT MATCH|      |
|NUMBER   SCORE   COUNT   MEASURE   S.E. |MNSQ    ZSTD|MNSQ    ZSTD|CORR.   EXP.| OBS%   EXP%| ITEM |
|----------------------------------------+-----------+------------+------------+-----------+------|
|      7       3      20      3.74    .82|1.14     .44|1.46     .76|  .47    .57| 89.5   89.2| Q17  |
|      9       8      20      1.54    .58| .63  -1.30| .56    -.94|  .77    .62| 94.7   78.4| Q19  |
|     10       9      20      1.21    .57| .63  -1.40| .52   -1.22|  .77    .61| 89.5   76.9| Q20  |
|      6      11      20       .58    .56|1.57    1.95|1.48    1.11|  .34    .58| 47.4   74.7| Q16  |
|      8      12      20       .28    .56| .68  -1.35| .52   -1.07|  .71    .55| 78.9   74.9| Q18  |
|      4      13      20      -.03    .56|1.30    1.20|2.32    1.94|  .34    .53| 73.7   74.6| Q14  |
|      1      17      20     -1.49    .69| .84    -.33| .67     .08|  .42    .36| 84.2   84.1| Q11  |
|      2      17      20     -1.49    .69| .91    -.12| .55    -.06|  .42    .36| 84.2   84.1| Q12  |
|      5      17      20     -1.49    .69|1.06     .28| .62     .04|  .37    .36| 84.2   84.1| Q15  |
|      3      19      20     -2.85   1.05|1.11     .41| .91     .40|  .14    .21| 94.7   94.7| Q13  |
|----------------------------------------+-----------+------------+------------+-----------+------|
| MEAN     12.6    20.0       .00     .68| .99      .0| .96      .1|             | 82.1   81.6|      |
| P.SD      4.8      .0      1.82     .15| .29     1.1| .57     1.0|             | 13.1    6.5|      |
```

3.2. Item and Person Reliability

Reliability means whether an instrument can be interpreted consistently across different situations (Field, 2009). The common internal consistency parameters used are the Cronbach Alpha coefficient (Taherdoost, 2016). The Cronbach Alpha coefficient is the interaction between a person's ability and item difficulty. RMM can provide index values of two types of reliability, namely the Person Reliability Index (PRI) and the Item Reliability Index (IRI). The PRI shows the replicability of students that we can predict if this sample of person is given to a set of parallel items that measure the same construct (Bond & Fox, 2015).

Table 3 shows that PRI was 0.55. This indicates that the consistency of answers from respondents is pretty good (Linacre, 2019). Table 3 also shows the average values of infit and outfit of ZSTD were -0.03 and 0.19, respectively. These values mean that the quality of the person is getting better. The Person Separation Index (PSI) was 1.09. The PSI shows the number of strata of individual ability identified in the group studied. A low PSI indicates that the instrument may not be sensitive to differentiate between low and high achievers, thus, more items may be needed (Linacre, 2019).

Table 3. Person Reliability of MCHP

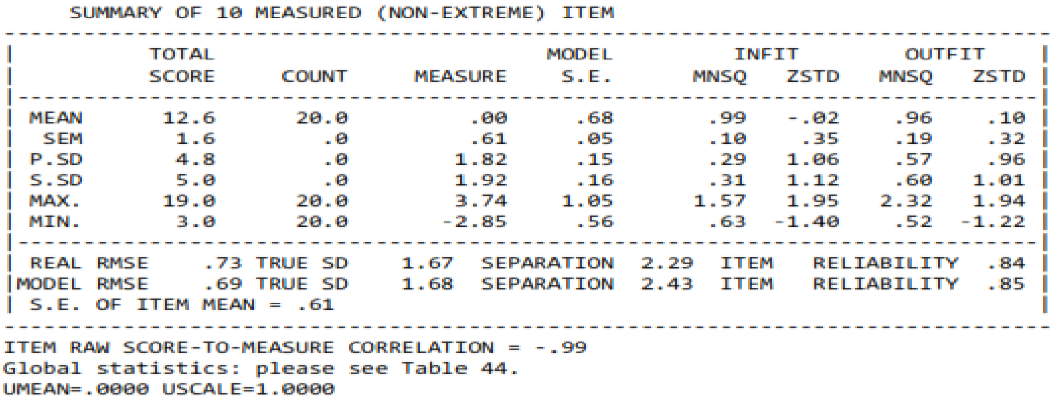


```
     SUMMARY OF 19 MEASURED (NON-EXTREME) PERSON
-------------------------------------------------------------------------------
|         TOTAL                          MODEL         INFIT        OUTFIT    |
|         SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       6.1       10.0         .81    .92       .96   -.03    .96    .19 |
|  SEM        .5         .0         .36    .04       .12    .21    .22    .18 |
| P.SD       2.0         .0        1.51    .17       .49    .91    .91    .74 |
| S.SD       2.0         .0        1.55    .17       .50    .93    .94    .76 |
| MAX.       9.0       10.0        3.33   1.27      2.17   2.24   3.07   1.83 |
| MIN.       3.0       10.0       -1.37    .80       .32  -1.25    .11   -.63 |
|-----------------------------------------------------------------------------|
| REAL RMSE    1.02 TRUE SD    1.11  SEPARATION  1.09  PERSON RELIABILITY  .55 |
|MODEL RMSE     .93 TRUE SD    1.19  SEPARATION  1.27  PERSON RELIABILITY  .62 |
| S.E. OF PERSON MEAN = .36                                                    |
-------------------------------------------------------------------------------

PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .67  SEM = 1.21
```

The IRI indicates the replicability of item assignment along the track if these same items are provided to another same size sample of students who have the same behavior (Bond & Fox, 2015). Table 4 shows that IRI was 0.84. According to the qualitative descriptor used for Cronbach Alpha value, the item reliability of MCHP is reliable and acceptable (Taber, 2018). Table 4 also shows the average values of infit and outfit of ZSTD are -0.02 and 0.19, respectively. This means that the quality of the item is getting better. The item separation index (ISI) was 2.29. This index serves to verify the hierarchy of items. A low ISI (below 3) indicates that the person sample is not large enough to confirm the item difficulty of the instrument. Thus, more persons may be needed (Linacre, 2019).

Table 4. Item Reliability of MCHP

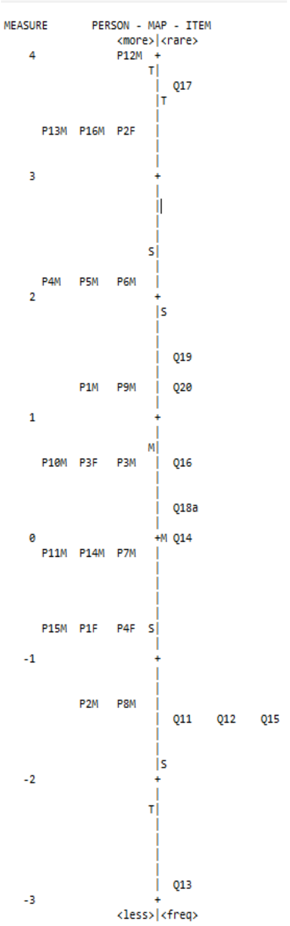


```
     SUMMARY OF 10 MEASURED (NON-EXTREME) ITEM
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       12.6       20.0         .00    .68       .99   -.02    .96    .10 |
|  SEM        1.6         .0         .61    .05       .10    .35    .19    .32 |
| P.SD        4.8         .0        1.82    .15       .29   1.06    .57    .96 |
| S.SD        5.0         .0        1.92    .16       .31   1.12    .60   1.01 |
| MAX.       19.0       20.0        3.74   1.05      1.57   1.95   2.32   1.94 |
| MIN.        3.0       20.0       -2.85    .56       .63  -1.40    .52  -1.22 |
|-----------------------------------------------------------------------------|
| REAL RMSE     .73 TRUE SD    1.67  SEPARATION  2.29  ITEM   RELIABILITY  .84 |
|MODEL RMSE     .69 TRUE SD    1.68  SEPARATION  2.43  ITEM   RELIABILITY  .85 |
| S.E. OF ITEM MEAN = .61                                                      |
-------------------------------------------------------------------------------
ITEM RAW SCORE-TO-MEASURE CORRELATION = -.99
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

3.3. Item-Person Map

The item-person map illustrates the linear relationship between the level of difficulty of the items and the ability of the participants. Figure 2 presents a map of people's abilities, item difficulties, and the appropriate distribution of item difficulties and student competence. The most difficult items are displayed at the top of the image to the right of the y-axis and the most capable students are the highest on the left side. If the score participant is plotted at the same level as the item, then the participant has a 50% chance of answering the item correctly (Linacre, 2019). If the item level is above the participant, then the possibility of the participant answering the item correctly is less than 50%. This is due to the item being approximated to be outside their ability level.

The most difficult item was question 17 (Q17) located at the top of the map, while the easiest item was question number 13 (Q13) at the bottom of the map as shown in Figure 1. Q11, Q12, Q13, and Q15 items were located below all persons on the map. This indicates that all students can answer these questions. Mokshein, et al. (2019) suggested that these items should be revised because they cannot differentiate between student abilities. This can occur because the test is carried out on a small size sample. These items are expected to differentiate student abilities if the trial is conducted on a big-size sample. Therefore, the researcher wants to retain these items for testing against a large sample because they are categorized as valid items.



```
MEASURE       PERSON - MAP - ITEM
                  <more>|<rare>
    4             P12M  +
                       T|
                        |  Q17
                        |T
                        |
      P13M  P16M  P2F   |
                        |
                        |
    3                   +
                        |
                        ||
                        |
                        |
                       S|
                        |
      P4M   P5M   P6M   |
    2                   +
                        |S
                        |
                        |
                        |  Q19
                        |
            P1M   P9M   |  Q20
                        |
    1                   +
                        |
                       M|
      P10M  P3F   P3M   |  Q16
                        |
                        |
                        |  Q18a
                        |
    0                   +M Q14
      P11M  P14M  P7M   |
                        |
                        |
                        |
                        |
      P15M  P1F   P4F  S|
                        |
   -1                   +
                        |
                        |
            P2M   P8M   |
                        |  Q11    Q12    Q15
                        |
                        |
                        |S
   -2                   +
                        |
                       T|
                        |
                        |
                        |
                        |
                        |  Q13
   -3                   +
                  <less>|<freq>
```

Figure 2. Item-Person Map

3.4. Analysis of Extreme Items

Analysis of extreme items needs to be conducted to observe the weaknesses of the items and fix them for further development. Based on the Rasch Model Measurements (RMM) made, it was found that Q13 was the easiest item while Q17 items were the most difficult.

3.4.1. Analysis of easiest item

Figure 3 shows the easiest question, Q13. Q13 aims to test the competence of students in the concept and calculation of the ideal gas law and its application in everyday life. Q13 provides information about determining the relative molecular mass (M_r) of a compound by heating a volatile substance with a certain mass and measuring the volume and pressure of the gas produced at a certain temperature.

The relative molecular mass of a volatile liquid can be determined by heating the liquid in water at a temperature of 100^0C and measuring the volume of gas collected in the injection (as in Figure A). A total of 0.2245 grams of the volatile liquid sample is heated to evaporate. If the evaporation results can fill 81 mL at 1 atm and 100^oC, then the relative molecular mass of the liquid is…

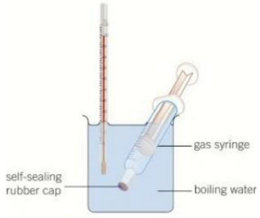


Figure A. The molecular mass determination apparatus for volatile compounds

Figure 3. Question 13 (Q13)

Students were asked to find the relative molecular mass of a compound based on the information in the problem. Almost all students (95%) were able to answer this question correctly. Students were able to calculate the relative molecular mass of compounds using the ideal gas equation correctly. This occurs because students are provided with pictures and methods to determine the relative molecular mass with an ideal gas equation. According to (Da Silva & De Castro, 2018), the use of simple equipment handled by students is an effective method for learning science concepts.

3.4.2. Analysis of most difficult item

Figure 3 shows the most difficult question, Q17. Q17 aims to test the ability of students about the concepts and calculation of moles and percentage yield and their application in daily life. Q17 provides information about making aspirin from salicylic acid by acetylation reaction with a certain percentage yield. Students were asked to determine the mass of salicylic acid needed to produce a certain number of aspirin. Almost all students (85%) were unable to answer this question correctly. Only a few students (15%) were able to complete Q17 well. These students were at the top left on the item-person map. The analysis of Q17 indicates that almost all students cannot understand the "percentage yield" and how to use them in stoichiometry problems. Most (about 60%) could not show an adequate understanding of how to use percentage yield for certain reaction products. They mostly multiply the mass of the reaction product by a percentage number and divide it by one hundred. They should first multiply the mass of the reaction product by

100 and divide it by the percentage of yield. These results are consistent with the research results by Hanson (2016), which find out that the majority of prospective chemistry teachers in Ghana do not understand using percentage results in chemical calculations. A few students (20%) failed to solve this question with a blank in the worksheet. This can occur because students do not understand the meaning of the questions well. HOTS problems have several words and sentences that students are difficult to understand and solve the problems (Sa'adah et al., 2019). A few students (15%) laid in the left-top item-person map could solve this problem correctly.

Aspirin, $C_9H_8O_4$, is prepared from the acetylation reaction of salicylic acid, $C_7H_6O_3$, according to the equation:
$C_7H_6O_3 + (CH_3CO)_2O \rightarrow C_9H_8O_4 + CH_3COOH$
If the percentage yield for this reaction is 83%, then the mass of salicylic acid needed to make 1 kg of aspirin is…. (A_r C = 12, H = 1, O = 16)

Figure 4. Question 17 (Q17)

4. CONCLUSION

MCHP showed there was no negative correlation point measure and had one item that was not appropriate. Almost all items were valid, except for the Q16 items. Obtained 9 valid items and could be used for a larger sample for further study. RMM suggested adding the number of items and sample students for the next analysis. RMM provided useful information that could be used for the next development assessments and to establish a pedagogical assessment validation process. This research was at an early stage in the developing HOTS problems on stoichiometry. This study was limited to testing HOTS problems in a small sample (less than 30). Further research must develop more HOTS questions and conduct tests on larger samples (more than 100 students) to increase the reliability and validity of the developed HOTS questions.

REFERENCES

Abdullah, Rasmiwetti, & Ardiansyah. (2019). Profile of Superior High School Students' Higher Order Thinking Skills on Stoichiometry Topic. *Journal of Educational Sciences*, *3*(2), 174–183.

Bond, T. G. (2010). Rasch Model. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (4th ed., p. 773). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479216.corpsy0773

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences* (3rd ed.). Routledge. https://doi.org/10.1017/CBO9781107415324.004

Cavanagh, R. F., & Waugh, R. F. (2011). Applications of Rasch Measurement in Learning Environments Research. In *Sense Publisher* (Vol. 2). Sense Publisher. https://doi.org/10.1017/CBO9781107415324.004

Cola, N., & Nuswowati, M. (2020). Improvement Of Critical Thinking Ability In Acid Base Material Through The Application Of Edmodo-Based Discovery Learning Model. *JTK (Jurnal Tadris Kimiya)*, *1*(1), 61–70.

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., Karaduman, A., Swan, K., Clavé, P., & Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, *33*(4), 441–456. https://doi.org/10.1007/s00455-017-9873-4

Da Silva, L. M., & De Castro, A. H. (2018). Revisiting ideal gases and proposal of a simple experiment for determining atmospheric pressure in the laboratory. *Quimica Nova*, *41*(7), 818–824. https://doi.org/10.21577/0100-4042.20170221

Dahlan, D., Permana, L., & Oktariani, M. (2020). Teacher's competence and difficulties in constructing hots instruments in economics subject. *Cakrawala Pendidikan*, *39*(1), 111–119. https://doi.org/10.21831/cp.v39i1.28869

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd editio). Sage Publication.

Habibi, H., Jumadi, J., & Mundilarto, M. (2019). The rasch-rating scale model to identify learning difficulties of physics students based on self-regulation skills. *International Journal of Evaluation and Research in Education*, *8*(4), 659–665. https://doi.org/10.11591/ijere.v8i4.20292

Hanson, R. (2016). Ghanaian Teacher Trainees' Conceptual Understanding of Stoichiometry. *Journal of Education and E-Learning Research*, *3*(1), 1–8. https://doi.org/10.20448/journal.509/2016.3.1/509.1.1.8

Jailani, J., Sugiman, S., & Apino, E. (2017). Implementing the problem-based learning in order to improve the students' HOTS and characters. *Jurnal Riset Pendidikan Matematika*, *4*(2), 247–259. https://doi.org/10.21831/jrpm.v4i2.17674

Karlin, O., & Karlin, S. (2018). Making Better Tests with the Rasch Measurement Model. *InSight: A Journal of Scholarly Teaching*, *13*(1), 76–100. https://doi.org/10.46504/14201805ka

Kurniawan, O., & Noviana, E. (2017). Penerapan Kurikulum 2013 Dalam Meningkatkan Keterampilan, Sikap, Dan Pengetahuan. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan Dan Ilmu Pendidikan Universitas Riau*, *6*(2), 389–396.

Kusuma, M. D., Rosidin, U., Abdurrahman, A., & Suyatna, A. (2017). The Development of Higher Order Thinking Skill (Hots) Instrument Assessment In Physics Study. *IOSR Journal of Research & Method in Education (IOSRJRME)*, *07*(01), 26–32. https://doi.org/10.9790/7388-0701052632

Linacre, J. M. (2019). *A User's Guide To WINSTEP MINISTEP Rasch-Model Computer Programs*. https://doi.org/ISBN 0-941938-03-4

Lisdiani, S. A. S., Setiawan, A., Suhandi, A., Malik, A., Sapriadi, & Safitri, D. (2019). The Implementation of HOT Lab Activity to Improve Students Critical Thinking Skills. *Journal of Physics: Conference Series*, *1204*, 1–5. https://doi.org/10.1088/1742-6596/1204/1/012033

Misrom, N. S., Abdurrahman, M. S., Abdullah, A. H., Osman, S., Hamzah, M. H., & Fauzan, A. (2020). Enhancing Students ' Higher -Order Thinking Skills (HOTS) Through an Inductive Reasoning Strategy Using Geogebra. *International Journal of Emerging Technologies in Learning*, *15*(3), 156–179.

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, *204*(November 2014), 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. *Cakrawala Pendidikan*, *38*(1), 16–32. https://doi.org/10.21831/cp.v38i1.22750

Ng, S. E., Yeo, K. J., & Mohd Kosnin, A. B. (2018). Item Analysis for the Adapted Motivation Scale Using Rasch Model. *International Journal of Evaluation and Research in Education (IJERE)*, *7*(4), 264–269. https://doi.org/10.11591/ijere.v7i4.15376

Ramdiah, S., Abidinsyah, Royani, M., & Husamah. (2019). Understanding, planning, and implementation of HOTS by senior high school biology teachers in Banjarmasin-Indonesia. *International Journal of Instruction*, *12*(1), 425–440. https://doi.org/10.29333/iji.2019.12128a

Razali, S. N., Shahbodin, F., Ahmad, M. H., & Noor, H. A. M. (2016). Measuring Validity and Reliability of Perception of Online Collaborative Learning Questionnaire Using Rasch Model. *International Journal on Advanced Science Engineering Information Technology*, *6*(6), 966–974. https://doi.org/10.1109/IUSER.2016.7857960

Sa'adah, S. I., Rasmiwetti, R., & Linda, R. (2019). Pengembangan Soal Hots Dengan Wondershare Quiz Creator Sebagai Media Display Pada Materi Stoikiometri Kelas X. *JTK (Jurnal Tadris Kimiya)*, *4*(2), 177–188. https://doi.org/10.15575/jtk.v4i2.5469

Saido, G. M., Siraj, S., Bakar, A., Nordin, B., & Saadallah, O. (2015). Higher Order Thinking Skills among Secondary School Students in Science Learning. *Malaysian Online Journal of Educational Sciences*, *3*(3), 13–20.

Sarif, Y., Fadiawati, N., & Syamsuri, M. M. F. (2019). The Effectiveness of Waste Cooking Oil Recycling Project-Based Learning t o Improve Students ' High Order. *International Journal of Chemistry Education Research*, *3*(1), 29–34. https://doi.org/10.20885/ijcer.vol3.iss1.art

Susongko, P. (2016). Validation of science achievement test with the Rasch model. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 268–277. https://doi.org/10.15294/jpii.v5i2.7690

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), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*, *5*(3), 28–36. https://doi.org/10.2139/ssrn.3205040

Tan, S. Y., & Halili, S. H. (2015). Effective Teaching of Higher-Order Thinking (HOT) in Education. *The Online Journal of Distance Education and E-Learning*, *3*(2), 41–47.

Tyas, M. A., Nurkamto, J., Marmanto, S., & Laksani, H. (2019). Developing Higher Order Thinking Skills (HOTS) – Based Questions: Indonesian EFL Teachers' Challenges. *Developing Higher Order Thinking Skills (HOTS) – Based Questions: Indonesian EFL Teachers' Challenges*, *2*(1), 52–63. https://doi.org/10.17501/26307413.2019.2106

Wright, B. D., & Stone, M. (1999). *Measurement Essesntials* (2nd ed.). Wide Range Inc.