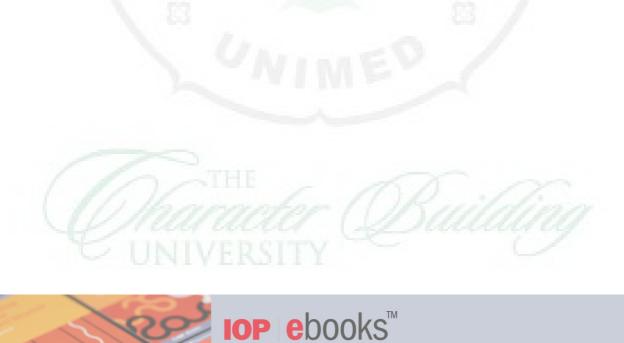**PAPER • OPEN ACCESS**

Development of The Thermochemistry- Hots-Tawheed Multiple Choice Instrument

View the article online for updates and enhancements.

Development of The Thermochemistry- Hots-Tawheed Multiple Choice Instrument

A Darmana[1*], A Sutiani[2], Jasmidi[2]

[1,2]Chemistry Department, Faculty of Mathematics and Natural Science, Universitas Negeri Medan, Sumatera Utara, Indonesia

Corresponding author email : ayidarmana2013@gmail.com

Abstract. This research is part of the development research in thermochemistry-hots-tawheed teaching materials. the purpose of this research is to develop the thermochemistry tawheed High Order Thinking skills (HOTs) test instrument. the stages of the research began by designing 45 multiple choice items with five options. The instrument assessed by three science education experts, and pilot tested to 56 chemistry students. analysis of the test items was carried out using the rasch measurement model approach, the multiple choice test analyzed with winsteps software, and expert validation analyzed using facets software. Multirater validation test results showed 27 of 45 test items were in good quality (60%) according to the experts. The multiple choice items results : (1) the reliability of the test items was good quality (reliability index 0.89); (2) test of item fit, 34 out of 45 test items are good in fit statistics indices (75.6%); (3) person map item test, showing difficulty level, easy: medium: difficult = 50%: 33%: 17%; (4) the test distractor work well for 23 of the 45 test items (51%); (5) there are 19 of 45 valid test items based on pilot tested results that are in accordance with expert judgment.

1. Introduction

The formation of a perfect human being as stated in the goal of National Education   meaning that Indonesian education must be able to produce Indonesian human resources which have a high level of cognitive ability and at the same time have *iman* and *taqwa* to Allah. High-level cognitive abilities are characterized as intelligent, creative, innovative human resources, able to collaborate and adapt to their environment, able to conduct analyzes in problem solving and decision making efforts. High-level cognitive abilities contribute to humans in running the role as caliph. While the *iman* to Allah   will contribute to the character that keeping   the cognitive ability to be a means of happiness and benefit of humanity, as an expression of gratitude to Allah who has given the ability to humans to think and at the same time provide a means which is the object of study.

The Program for International Students Assessment, PISA, with 72 countries participating, has reported that Indonesia until 2015 was still ranked 62nd in scientific ability [1]. On the other hand, the morality aspect that reflects the social attitudes, *iman* and *taqwa* are   increasing bad. Many cases are not commendable occur among students, fight, sexual harassment, the use of illegal drugs and free sex that all almost often witnessed in social media.

Looking at such conditions, the role of learning must simultaneously contribute to dealing with these two things. General chemistry courses as part of science and the curriculum of higher education must contribute in addition to increasing the ability to think at a higher level, must also be able to be a means of inculcating positive values. General chemistry learning must be able to initiate the formation

of a noble person as a reflection of the attitude of *iman* and *taqwa* to Allah. The ability to think at a high level can be equipped with HOTS learning, while for the inculcation of positive values can be done by including tawheed values in learning. Chemistry is very closely related to natural phenomena which are nothing but a form of the actualization of Allah's Greatness and Unity. The value of tawheed is the most important thing that shows the Unity of Allah.

Many studies have been conducted as a response to the above problem. This is shown by the many studies that focus on HOTS [2][3][4][5], as well as on social and spiritual attitudes [6][7][8][9][10][11][12], them selves have conducted several studies from 2014 to 2016 that focus on efforts to form attitudes through the integration of tawheed values in teaching material.The topic of thermochemistry is very full of natural phenomena that occur in our daily environment. Natural phenomena is about of what Allah has determined. Presenting religion in science will not reduce the scientific level of science and will even make the meaning of science itself [13]. Tawheed is the core of religion which means to proclaim Allah as the Creator of the universe, the Sustainer, the Owner, the Maker of the Decree and the right to be worshiped. Implanting the value of tawheed means instilling in students that everything studied basically belongs to Allah, all phenomena, Allah's laws that exist in nature which are basically the objects of scientific study. Events in nature.phenomena that occur in our environment basically occur with certain patterns, making it easier to learn, understand and apply to other situations according to human needs. Tawheed values will appear in the form of narration as a stimulus in the item Hots. Stimulus is usually presented in the form of diagrams, tables and descriptions of case / narrative fragments [4].

Based on these reasons, this research will examine the development of instruments to measure the ability of students in high-level thinking (HOTS) and at the same time be expected to be a means for the formation of the *iman* and *taqwa*. Integration of the value of tawheed (INT) will appear as a stimulus, so that the stimulus nuances reminiscent of the Power, Greatness and Unity of Allah. The study will emphasize to obtain a good multiple-choice test items based on expert judgment and trials

2. Research Method

The research design used for this study was adapted from research and development from Gall & Borg [14], The design includes four stages, namely a preliminary study, design, development and testing. In the preliminary stage, teaching materials, sets of equipmen of general chemistry learning especially thermochemistry are examined. As well the study of thermochemical concepts that make it possible to make instruments in the form of Hots and tawheed. The next step is to design the thermochemical - Hot-Tawheedinstrumen.The next step is expert validation and testing. Expert validation involved three experts (experts on the preparation of Hots, General chemists and physical chemists, and general chemists and learning evaluations). A trial was conducted on 56 students of the chemical education study program class of 2018 (semester 3).

The instruments used in this study were two instrumen. The first instrument, the validation instrument of the Thermochemical-Hots-Tawheed PG test instrument used by expert validators. This instrument was adapted from the "examine hots multiple choice test" instrument in a book on how to make of hots tes item [15]. The second instrument is an instrument to measure the ability of thermochemical-Hots-tawheed consisting of 45 multiple choice items with five options. The data have collected is instrument characteristic data based on expert judgment (validator) and trial. Expert assessment data is a multirater validator data which includes validity of the thermochemical-Hots-Tawheed PG test items, data on the level of difficulty of the validation instrument statement items answered by the validator, as well as data on the level ofeasy (not stingy) of the validator in providing an assessment. The trial data includes data on item reliability, item person maps, item fit statistics, and distractor analysis. The data collected was then analyzed using the Rasch model objective measurement approach with Facets software for expert validation testing and Winsteps software for test problems [16].

3. Result and Discussion

3.1. The result of expert judgment (multiratervalidation)

Description of multiater validation showed in Tabel 1

Table 1. Results of multirater validation tests with facets

```
+-------------------------------------------------------------+------------------+-------+
|Measr|+ButirSoal                                             |-Aitem            |-Valida|
|----+------------------------------------------------------+------------------+-------|
|  3 +                                                        +                  +       |
|    |                                                        |                  |       |
|    |                                                        |  M3              |       |
|    |                                                        |                  |       |
|  2 +                                                        +                  +       |
|    | S11  S25  S28  S5                                      |  M4              |       |
|    |                                                        |  M8              |       |
|    |                                                        |                  |       |
|  1 + S10  S13  S15  S16  S24  S29  S4   S40  S9             + K1  M5           +       |
|    |                                                        |                  |       |
|    | S17  S18  S2   S20  S26  S27  S32  S39  S44  S6   S7   | M2               |       |
|    |                                                        | K3               |       |
|*   0 * S14  S30  S8                                         * K2  K4  K7        *       *
|    | S45                                                    | M6               |       |
|    | S1   S12  S19  S38  S43                                |                  |       |
|    | S21  S22  S23  S42                                     | B1  B2  B3  B4  M7|       |
| -1 + S35                                                    + K8               +       |
|    | S34                                                    |                  |       |
|    | S31  S36  S37                                          |                  |       |
|    | S41                                                    | K5  K6           |       |
| -2 + S3                                                     +                  +       |
|    |                                                        |                  | DrZ  |
|    |                                                        |                  |       |
|    | S33                                                    |                  |       |
| -3 +                                                        +                  +       |
|    |                                                        |                  |       |
|    |                                                        |                  | DrR  |
|    |                                                        |                  |       |
| -4 +                                                        +                  +       |
|    |                                                        |                  |       |
|    |                                                        |                  | DrA  |
|    |                                                        |                  |       |
| -5 +                                                        + M1               +       |
|----+------------------------------------------------------+------------------+-------|
```

In table 1, a picture of multirater validity is presented. The image consists of 4 columns, namely the 1st column to the 4th column. The 1st column is the measure (logit transformation) column which states the measurement results in the price range of +3 (top) to -5 (bottom), this number is / is called the logit value. In column 2 (item analysis by experts), describe the distribution of items that are in the range of logit values from the smallest to close to -3 logit (item S33) to the largest value approaching +2 logit (item S11, S25, S28 and S5). A logit price of 0 is a minimum criterion of the quality of questions considered to be of quality by experts, so if the value is positive (greater than zero logit) indicates questions considered good by the expert panel, whereas if the value of the logits is negative (less than 0) indicates questions the problem is not good according to experts.

3.2. The validity of PG test items

From the picture above provides information that items that are considered not good by experts starting from item number 45 (S45), 1, 12, 19, 38, 43, 21, 22, 23, 42, 35, 34, 31, 36 , 37, 41, 3 and 33 (there are 18 items, 40%). While the items that are considered qualified by the expert panel are item 14, 30, 8, 17,18, 2, 20, 26, 27,32, 39,44,6,7,10,13, 15,16,24 , 29,4,40,9, 11,25, 28, and 5 (there are 27 valid items, 60%).

3.3. Distribution of the difficulty level of the validation instrument items

Information on the distribution of the difficulty level of the validation instrument statement items is shown in column 3 (item column) of Figure 1 above. The validator's response to the validation instrument statement items is the validator's assessment of the validation instrument statement items (item column). The column shows the distribution of material aspects (M), construction (K), and Language (B). M1 means the material aspects reviewed by the validator for the 1st statement (there are 8 statements for the material aspects, meaning there are M1 - M8). The same thing for K, K1

means the construction aspect which is reviewed by the validator for the 1st statement (there are 8 statements for the construction aspect, meaning there are K1 - K8). B1 means the language aspect reviewed by the validator for the 1st statement (there are 4 statements for the language aspect, meaning there is B1 - B4).

Validation instrument statement items that are analyzed with the lowest / lowest / negative negativelogit price indicate the validation instrument statement items that are most easily fulfilled by items according to the validator, whereas for positive logit prices, the more positive / greater / more the above shows the statement items which is difficult to find in the items assessed by the expert panel. For item evaluation item items that have the same logit price means the same level of ease / difficulty. From the item column picture 1 above, information is obtained that M3 (logit price is greater / greater than +2) is the most difficult / difficult to fulfill statement item from items evaluated according to the three validators. Whereas M1 (lowest logit price / smaller than -4) is the easiest item to fulfill from the item according to the three validators. M1 is a statement about "Questions in accordance with indicators", this shows that almost all the items that exist meet this aspect of assessment. Whereas M3 (with a value of more than +2 logits) is a statement about "The question of having a contextual-spiritual stimulus" shows that all items according to the three expert panels have not fulfilled this aspect.

3.4. The level of easy / generosity of the validator

In column 4 (validator) the distribution of the three validators is at a negative logit value (-2 to - 4) this provides information that the three validators, including assessors who are not so strict, give an assessment on the items given. The sequence of the most stringent judges how to grade it to the cheapest rate is Dr. Z. Dr. R. and Dr. A.

Table 2. the results of the Expert Panel validation test with Facets

```
+-------------------------------------------------------------------+
|              |             Model | Infit      Outfit   |Correlation |
|   Validator  |Measure   S.E. | MnSq  ZStd  MnSq  ZStd |PtMea      |
|--------------+---------------+------------------------+-----------|
|    Dr Z      |  -2.19    .11 |  .94   -.9   .68  -2.1 |  .50      |
|    Dr R      |  -3.52    .15 | 1.00    .0   .60  -1.3 |  .36      |
|    Dr A      |  -4.44    .20 | 1.22   1.3  1.52   1.0 |  .16      |
|--------------+---------------+------------------------+-----------+
Separation 5.75  Strata 8.00  Reliability (not inter-rater) .97
Model, Fixed (all same) chi-square:  118.9  d.f.: 2  significance (probability): .00
Inter-Rater agreement opportunities: 2700
Exact agreements: 2308 =  85.5%  Expected:  2332.0 =  86.4%
```

Table 2 explains the quality of the assessment by a panel of experts sorted by the level of strictness of the assessment (severity) of the item. From the aspect of statistical accuracy it appears that Dr R's experts are the most consistent (Meansquare Outfit and ZSTD Outfit values are both in the range). From the aspect of reliability among assessors, the value is very good (0.97); and the data according to the measurement model were shown to be significant from the Chi-square test value ($p < 0.01$). What is interesting is the similarity of the assessment by the three experts (exact agreement) which reached 85.5% which shows that there are not many differences in the three experts in evaluating all items.

3.5. The result of trail (pilot tests)

3.5.1. Reliability and model accuracy

The following is presented in Table 3, respondent test items and items

	Person	Item
N	56	44^
Measures		
Mean	-0.68	0.00
SD, standard devation	0.50	1.03

SE, standard error	0.02	0.08
Outfit Mean Square		
Mean	1.03	1.03
SD	0.34	0.19
Separation	0.97	2.86
Reliability	0.48	0.89
Alpha cronbach	0.49	
Chi-square (χ^2)	2721.41*	
Unidimensionality	20.4%	

^ ada 1 soal yang outlier, tidak ada satupun siswa yang mampu menjawabnya dengan benar

   * $p < 0.01$

The logit person average is rather low, -0.68 shows that the average ability level of participants is slightly below the level of difficulty of the questions (there are some questions that can be too difficult for some participants); with a standard deviation value of 0.5 indicating the range of achievement of test takers is not so large. Whereas for items with an average logit of 0.0 and an SD score of 1.03, the level of difficulty of the questions varies in measuring the ability of test takers from those who are slow to learn (negative logit scores around -2 logit) to high-ability test participants (logit value +2).

A low standard error value indicates meticulous measurement quality for both person (0.02) and item (0.08). A small person reliability index (0.48) indicates a small diversity of respondents (which is normal because the number of test participants is only 56 people and the type of data is dichotomous; to show the stability of the data in a multiple choice test needs at least 150 people more), also separation value of 0.97 which when rounded into one, shows that the data analyzed is not very reliable (the value is not 2). The same thing can be seen from the small alpha cronbah value, 0.49 (minimum value of at least 0.67). However, the reliability index for items shows a better thing, from the aspect of item reliability which is 0.89 (more than 0.67) and the separation value if rounded is 3 (at least 3). This shows the reliability of the questions is good, but the test taker's ability is not so diverse. The accuracy of the data with the model (MNSQ outfit for persons and items) approaches the ideal value of 1.0, with a significant chi-square value, indicating that the data can be analyzed with the Racsh model. The level of unidimensionality that indicates the instrument has the ability to measure but not so good, just above 20%.

3.6. Item person Map (variable map or Wright map)

      Item person maps provide information about the distribution of test takers' ability levels and the distribution of test item difficulty levels presented in Figure 2
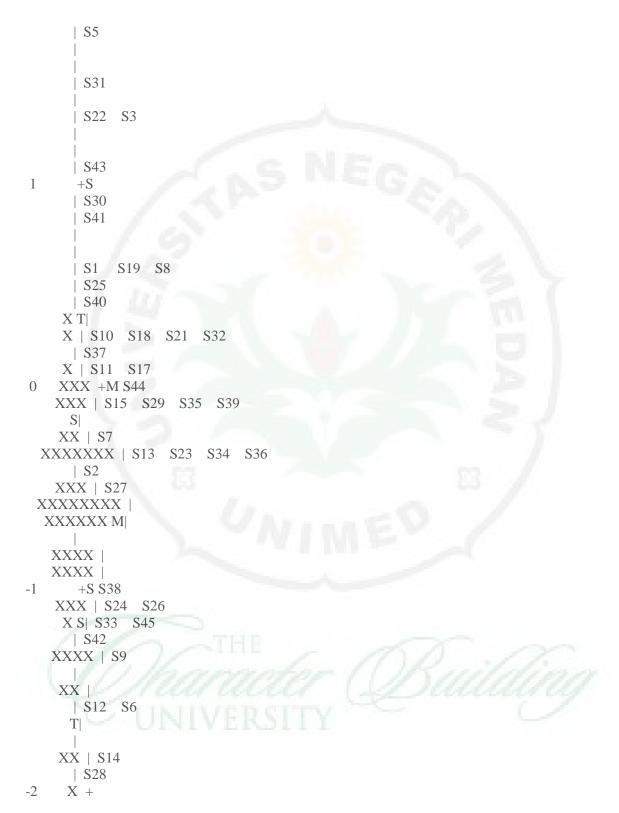
```
   Person - MAP - Item
<more>|<rare>
  3      +  S4
         |
         |
         | S16
         |
         |
         |
         |
         |
         |
         |
         |T
  2      +  S20
         |
         |
```

```
            | S5
            |
            |
            | S31
            |
            | S22   S3
            |
            |
            | S43
   1        +S
            | S30
            | S41
            |
            |
            | S1    S19   S8
            | S25
            | S40
        X T|
        X  | S10   S18   S21   S32
           | S37
        X  | S11   S17
   0    XXX +M S44
       XXX  | S15   S29   S35   S39
         S|
       XX  | S7
    XXXXXXX | S13   S23   S34   S36
          | S2
       XXX | S27
     XXXXXXXX |
      XXXXXX M|
           |
      XXXX |
      XXXX |
  -1        +S S38
      XXX  | S24   S26
       X S| S33   S45
          | S42
     XXXX | S9
          |
       XX |
          | S12   S6
        T|
          |
       XX | S14
          | S28
  -2      X +
```

Figure 2. distribution of respondens ability and tests dificulty

From picture 2 above, the left part of the diagram shows the distribution of the ability of the test takers (respondents), while the right part of the diagram shows the distribution of the level of test difficulty. The level of ability distribution of test takers has a logit price range between -2 (lowest) and the highest slightly above "0" (between 0 and 1). There are only 6 participants who have the ability with a

logit price greater than or equal to "0" which indicates the ability of the sufficient category. Whereas 50 more participants (56 test takers) had lower ability than the average level of difficulty of the questions (logit price below 0).

The level of difficulty of the questions is seen in the right part of the diagram, the more upward means the most difficult questions (example: S4 questions); while getting to the bottom is easier (for example: about S28). The following is presented in table 4 test item difficulty levels.

Tabel 4 . level of test item difficulcity

	The level of item difficulty				
	Very difficult	**Difficult**	**Moderate**	**Easy**	**Very easy**
Kode Soal	S4, S16	S20, S5, S31 S3, S22, S43	S30, S41, S1, S8, S19, S25, S40, S10, S18 S21, S32, S37 S11, S17, S44	S15, S29, S35 S39, S7, S13 S23, S34, S36 S2, S27, S38	S24, S26, S33 S45, S42, S9 S6, S12, S14 S28
Jumlah (persen)	2 (4%)	6 (13%)	15 (33%)	12 (27%)	10 (23%)

The item difficulty level classification is based on the values in Table 3 above, which is an average logit of 0.0 and a standard deviation of 1.03 logit. It can be seen that the level of difficulty of the item 50% is in the easy (easy and easy) category, 33% is moderate (15 questions), and in the difficult category is 17% (8 questions).Good question categories for diagnostic tests are: 50:30:20 each for the easy category: medium: difficult, the results of the analysis show that it is similar.

3.7. Test Item Fit

The following is the item fit test results presented in Table 5, Item fit tests are used to determine whether the test items function in measuring the ability to be measured (measurement accuracy)

Tabel 5. Item fit test

No	Item	Logit	Std Error	Outfit MNSQ	Outfit ZSTD	Point Meas Corr.
1	S1	0.58	0.32	0.99	0.00	0.23
2	S2	-0.45	0.28	0.93	-0.89	0.34
3	S3	1.35	0.41	1.01	0.15	0.14
4	S4	4.65	1.82	1.00	0.00	0.00
5	S5	1.74	0.47	0.92	-0.04	0.16
6	S6	-1.56	0.3	0.99	-0.04	0.27
7	S7	-0.22	0.28	0.82	-1.89	0.51
8	S8	0.58	0.32	0.84	-0.77	0.36
9	S9	-1.3	0.29	1.04	0.40	0.15
10	S10	0.29	0.3	1.22	1.35	-0.05
11	S11	0.11	0.29	0.99	-0.05	0.20
12	S12	-1.56	0.3	0.83	-1.18	0.50
13	S13	-0.37	0.28	1.03	0.37	0.15
14	S14	-1.84	0.32	1.21	1.12	-0.10
15	S15	-0.06	0.29	1.04	0.39	0.13
16	S16	2.72	0.72	1.93	1.22	-0.07
17	S17	0.11	0.29	0.85	-1.13	0.42
18	S18	0.29	0.3	1.14	0.89	-0.03
19	S19	0.58	0.32	1.14	0.73	-0.04
20	S20	1.98	0.52	1.12	0.39	-0.01

21	S21	0.29	0.3	0.96	-0.19	0.26
22	S22	1.35	0.41	1.08	0.35	0.02
23	S23	-0.37	0.28	1.03	0.42	0.15
24	S24	-1.06	0.28	0.86	-1.60	0.47
25	S25	0.48	0.31	1.23	1.20	-0.02
26	S26	-1.06	0.28	1.01	0.15	0.21
27	S27	-0.52	0.28	0.90	-1.39	0.40
28	S28	-1.94	0.32	0.91	-0.37	0.32
29	S29	-0.06	0.29	1.02	0.22	0.16
30	S30	0.92	0.35	1.03	0.20	0.26
31	S31	1.53	0.44	0.91	-0.11	0.18
32	S32	0.29	0.3	1.40	2.27	-0.17
33	S33	-1.14	0.28	1.01	0.14	0.19
34	S34	-0.37	0.28	0.99	-0.09	0.24
35	S35	-0.06	0.29	0.81	-1.68	0.52
36	S36	-0.37	0.28	1.05	0.66	0.15
37	S37	0.2	0.3	0.92	-0.48	0.32
38	S38	-0.98	0.28	0.90	-1.19	0.41
39	S39	-0.06	0.29	0.88	-1.05	0.41
40	S40	0.38	0.31	1.08	0.49	0.18
41	S41	0.8	0.34	1.28	1.13	-0.04
42	S42	-1.22	0.28	0.87	-1.29	0.44
43	S43	1.05	0.37	1.24	0.86	-0.05
44	S44	0.03	0.29	1.18	1.42	-0.05
45	S45	-1.14	0.28	0.88	-1.35	0.46

From the aspect of measurement accuracy (fit statistic), there is one problem that does not meet the MNSQ outfit criteria, namely item 16 (S16), the ideal value for MNSQ outfit is 0.5 - 1.5. For the ZSTD outfit criteria there is one problem with item 32 (S32), the ideal value is -2 to +2. While those who did not meet the criteria for Point measure correlation, where the value was negative, there were 11 items (S10, S14, S16, S18, S19, S20, S25, S32, S41, S43 and S44). It can be concluded that item S16 and S32, is a poor quality question; while the questions that need to be revised for improvement because of the potential to confuse students are S10, S14, S18, S19, S20, S25, S41, S43 and S44. While the questions that were declared valid statistically were 34 out of 45 question items (75.6%). The numbers are questions no: S1, S2, S3, S4, S5, S6, S7, S8, S9, S11, S12, S13, S15, S17, S21, S22, S23, S24, S26, S27, S28, S29, S30, S31, S33, S34, S35, S36, S37, S38, S39, S40, S42, and S45.

3.8. Test of distractor items

By analyzing the multiple choice questions rasch modeling can be identified whether distractors function or not. Whether or not the functioning of the supplier is known from the average logit score of the test taker who gets the correct answer, if the average logit person is smaller then that is an indication that the deception of the answer is not functioning. Analysis of the test results of deceitful items obtained item items that deceiters function well is item items no: S4, S5, S31, S3, S30, S43, S8, S40, S37, S17, S15, S35, S39, S7, S34, S27, S38, S26, S33, S45, S42, S9, and S12 (there were 23 out of 45 items or 51% percentage). While as many as 22 question items (49%) were stated to be misleading.

4. Conclusion

There are 27 out of 45 test items (60%) of validated multirater validation test results. The trial results show: (1) The reliability of the test items is stated to be good (reliability index 0.89); (2) Test item fit, 34 out of 45 test items are good in the statistical aspects (75.6%); (3) Person Map item test, showing difficulty level, easy: medium: difficult = 50%: 33%: 17%; (4) Deception works well for 23 of the 45

test items (51%); (5)There are 19 of 45 valid test items based on trial results that are in accordance with expert judgment.

References

[1]   Biro Komunikasi dan Layanan Masyarakat Kementerian Pendidikan dan Kebudayaan. https://www.kemdikbud.go.id/main/blog/2016/12/peringkat-dan-capaian-pisa-indonesia-mengalami-peningkatan. Diakses 14 Desember 2019

[2]   Anisah dan Lastuti,S.(2018, Desember).Pengembangan Bahan Ajar berbasis HOTS untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Mahasiswa. *Jurnal Matematika Kreatif-Inovatif,* 9(2), 191-197. http://journal.unnes.ac.id/nju/index.php/ kreano.

[3]   Kusuma, M.D., Rosidin, U., Abdurrahman, Suyatna, A. (2017, Februari), The Development of Higher Order Thinking Skill (HOTS) Instrument Assessment In Physics Study. *IOSR Journal of Research &Method in Education*, 7(1), 26-32. http://www.iosrjournals.org/iosr-jrme/papers/Vol-7%20Issue-1/Version-5/D0701052632.pdf

[4]   Ningsih, D.L. (2018). Analisis Soal Tipehigher Order Thingking Skill(Hots)Dalam Soal Ujian Nasional (UN) Biologi Sekolah Menengah Atas Tahun Ajaran 2016/2017. Skripsi. FKIP UNLA

[5]   Widana, I. (2017, February 27). Higher Order Thinking Skills Assessment (Hots). *Jisae: Journal Of Indonesian Student Assesment And Evaluation*, *3*(1), 32 - 44. https://doi.org/https://doi.org/10.21009/JISAE.031.04

[6]   Asysyifa, D.S., Sopyan, A., Masturi. (2017). Pengembangan Bahan Ajar Ipa Berbasis Komplementasi Ayat-Ayat Sains Quran Pada Pokok Bahasan Sistem Tata Surya. *Unnes Physics Education Journal.* 6 (1) . 44-54.

[7]   Darmana, A., Batubara, M.(2016). The Effect Of Thermochemistry Learning Based On Tawhed To Conceptual And Tawhed Understanding Of Senior High School Students *Proceedings of theAnnual International Seminar on Transformative Education and Educational Leadership (AISTEEL).335-340*

[8]   Faizah, N.S., Mubin, M. (2018). Pengaruh Modul Tematik Berbasis Integrasi Islam Dan Sains Pada Tema Energi Dan Perubahannya Terhadap Hasil Belajar Siswa Mi Murni Sunan Drajat Lamongan. *JPPIPA (Jurnal Penelitian Pendidikan IPA*). 3 (2). 1-5

[9]   Khairi, A., Agussuryani, Q., dan Hartini, P. (2017). Penumbuhan Karakter Islami melalui Pembelajaran Fisika Berbasis Integrasi Sains-Islam. *Jurnal Keguruan dan Ilmu Tarbiyah* . 02 (1). 19-31

[10] Kurniasih, W.(2016). Implementasi Integrasi Agama Dan Sains Dalam Pembelajaran Matematika Di Sekolah Dasar Islam Terpadu Harapan Umat Brebes, *Dialektika P. Matematika* . 3(2) ,20-29

[11] Permadi, B.A.(2018).Pengembangan Modul IPA Berbasis Integrasi Islam Dan Sains Untuk Meningkatkan Hasil Belajar Siswa Kelas Vi Min 2 Mojokerto, *Jurnal Pendidikan Islam*. 1 (2) . 294-311

[12] Wastyanti, A.(2016). Pengembangan Bahan Ajar Integrasi Sains Dengan Agama Berbasis Multimedia Di Madrasah Ibtidaiyah Negeri Malang 2. *Tesis, Pascasarjana Universitas Islam Negeri Maulana Malik Ibrahim Malang*

[13] Darmana, A. (2012). Internalisasi Nilai Tauhid dalam Pembelajaran Sains.Media pendidikan *: Jurnal pendidikan Islam, 27 (1), 66- 84*

[14] Borg,W. R. & Gall, M.D. (1983). *Educational researcher: An introduction*, (7th ed.). United States : Pearson education, Inc

[15] Sani, R.A. (2019). *Cara membuat soal hots*, Tangerang: Tira Smart

[16] Sumintono, B. & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH pada Assessment Pendidikan*, Cimahi : Trim Komunikata

.