PAEDAGOGIA
Jurnal Penelitian Pendidikan
http://jurnal.uns.ac.id/paedagogia

DOI: 10.20961/Paedagogia.V27i3.96615
Vol. 27, No. 3, 2024, pp. 470-480
p-ISSN 0126-4109
e-ISSN 2549-6670

Development of Five-Tier Multiple Choice (5TMC) Instrument to Analyze Misconceptions in Stoichiometry Material

Idha Ayu Kusumaningrum*, Maulidiningsih
Department of Chemistry Education, Faculty of Education and Teacher Training, Siber Syekh Nurjati Islamic University, Indonesia.

Keywords: *Stoichiometry, Misconception, Five-tier multiple-choice, Aiken, Rasch Model*

Article history
Received: 19 September 2024
Revised: 27 October 2024
Accepted: 27 October 2024
Published: 31 October 2024

*Corresponding Author Email: idhaayukusumaningrum@uinssc.ac.id

doi: 10.20961/paedagogia.v27i3.96615

Abstract: Stoichiometry is the basic of chemical calculations. However, students' understanding of the material is still poor. The characteristics of the material consist of abstract concepts, calculations, and related with other materials. That can cause misconception. The instrument to analyze misconceptions and the cause of student's misconceptions is still rare. This study aims to determine the feasibility of the five-tier multiple choice (5TMC) instrument and analyze misconceptions and the causes of misconceptions. The instrument was developed using the 4-D model (Define, Design, Develop, and Dissemination). This research was conducted with 103 respondents. Data was obtained from questionnaires, interviews, and tests. Content validity used the Aiken method with 5 experts. This research also used the Rasch Model to analyze the reliability and validity of the instrument. The research concluded that the five-tier multiple choice (5TMC) instrument developed was feasible and the criteria were good based on Aiken validity, Rasch model, and response questionnaire. Misconception was found in 2nd students as moderate (38,09%) and 4th students as low (24,94%). Personal thinking is the factor that most influences misconceptions. This instrument can be used as an alternative for analyzing student's misconceptions so that lecturers can take remediation.

How to cite: Kusumaningrum, I. A. & Maulidiningsih. (2024). Development of Five-Tier Multiple Choice (5TMC) Instrument to Analyze Misconceptions in Stoichiometry Material. *PAEDAGOGIA*, *27*(3), 470-480. doi: 10.20961/paedagogia.v27i3.96615

INTRODUCTION

The results of the questionnaire showed that 80% of chemistry students at UIN Siber stated that chemistry was difficult, especially in stoichiometry, thermodynamics, and chemical equilibrium. This opinion was supported by the results of interviews with lecturers who teach chemistry courses. They asked if stoichiometry material had the lowest score compared to other materials. Many students only memorized formulas without understanding. These findings are parallel to the research of (Woldeamanuel,M., Atagana,H., & Engida, 2014; Drastisianti et al., 2018; Mamombe et al., 2021). Characteristics of this material consist of many abstract concepts, calculations, and relating to other concepts (Opara, 2014). Chemistry consists of three levels; the macroscopic level refers to things that can be captured with the five senses. Second, the submicro level refers to symptoms at the molecular, atomic, ionic, and structural levels. Third, the representation level refers to symbols, formulas, equations, molarity, mathematical manipulations, and graphs (Johnstone, 2000). These three levels must be mastered by students to understand chemical concepts. Most students have difficulty, especially at the submicro level (Siswaningsih et al., 2019; Jammeh et al., 2023; Üce & Ceyhan, 2019). This concept cannot be observed directly with the eye, so the correct method is needed so that students can understand it. Inappropriate teaching methods will make students find it difficult to understand chemical concepts (Salame et al., 2020). This can also lead to misconceptions.

Misconceptions are situations when student's concept is different with scientists' concepts (Soeharto et al., 2019; Supatmi et al., 2019). However, students' misconceptions, especially in chemistry, rarely seem to be analyzed in learning activities. The lecturer focused on helping students achieve learning indicators without realizing if students may suffer misconceptions (Soeharto & Csapó, 2022). Misconception is one of the problems that disrupt the learning process and affect the following material because chemical materials are related to other (Mufit et al., 2018). Misconceptions can occur because students misunderstand the learning obtained, textbooks used by students, information sources from the internet, teacher's method, and students' daily experiences (Üce & Ceyhan, 2019; Härmälä-Braskén et al., 2020; Malaterre et al., 2023; Rahayu et al., 2024). Therefore, misconceptions need to be identified early so that they can be eliminated and not sustainable (Santi & Rahayu, 2022). Based on the results of the needs questionnaire, lecturers have never identified student's misconceptions. In fact, low student understanding of chemistry can occur because students have misconceptions. 90% of students stated that they need an instrument to analyze misconceptions and factors that cause them, especially in stoichiometry. Even though this is very important to do so that if there are misconceptions in students, remediation efforts can be made. Many factors cause student misconceptions, among others: student preconceptions from experience, teaching methods, materials, books, and solutions for a complete understanding called misconceptions that come from school (Barke H D, 2009). Misconceptions can also come from teachers who have misconceptions (Supatmi et al., 2019). Many misconceptions are found in chemical materials, especially in stoichiometry material (Lestari et al., 2021; Polamolo & Lukum, 2022; Rina Elvia et al., 2022; Sa'adah et al., 2022; Kristyowati & Priatmoko, 2023).

Stoichiometry material is the basis of chemical calculations. Almost all chemical calculations use the concept of stoichiometry. This causes students to understand the concept. If students do not understand/have misconceptions in this material, then in the next material, students will experience difficulties. Stoichiometry material consists of basic laws of chemistry, chemical reactions, and chemical calculations that require a lot of practice. Many studies show that students have trouble understanding in all of this material (Taha et al., 2014). This material requires more concept understanding rather than just relying on memorization. However, based on the results of interviews with lecturers teaching basic chemistry, students' understanding of stoichiometry material is still low. They only rely on memorizing formulas.

Many ways are used to identify misconceptions, one of which is through diagnostic tests. Diagnostic tests are an effective tool that identified the difference between what is the student learning and what we are expecting for their learning (Kanwal & Farooq, 2021). So, diagnostic test can help in identifying students who understand, do not understand, and misconception (Shiddiqi et al., 2024). Multiple choice is one example of a diagnostic test. Development of multiple choice on students' misconception gives contribution in misconception research (Gurel et al., 2015). Multiple choice tests continue to be developed from two-tier multiple choice to five-tier multiple choice. The five-tier multiple choice diagnostic test has the advantage of being able to diagnose misconceptions more deeply, and the source of the cause of misconceptions can also be found (Mardeni, 2023). This causes researchers to want to develop a five-tier multiple choice diagnostic test to identify misconceptions in chemical stoichiometry material so that if misconceptions are found, remedial learning can be held to correct the wrong concept.

METHOD

Research Design

This study was conducted to develop a five-tier multiple-choice test. The type of research was research and Development (R&D) with the 4-D model (Define, Design, Develop, and Disseminate) developed (Thiagarajan, S., Semmel & Semmel, 1976). Instrument development. The Define stage was conducted to collect information and analyze problems related to chemistry learning evaluation activities, which consisted of initial analysis, learner analysis, task analysis, concept analysis, and learning objectives. The design stage was carried out by making question grids and drafting questions that will be converted to Google Forms. The development stage was carried out with expert validation, small-scale trials, medium-scale trials, and wide-scale trials, followed by misconception analysis. The dissemination stage was carried out by presenting the research results to students and lecturers.

Research Target

Small-scale trials were carried out on 6 semester chemistry tadris students totaling 9 people, medium-scale trials were carried out on 2 semester biology tadris students totaling 46 people, and large-scale trials were carried out on 2 and 4 semester chemistry tadris students totaling 48 students. This research was conducted in the even semester of the 2023/2024 academic year.

Research Data

The instrument's feasibility was obtained from the results of expert validity, Rasch model analysis, and feasibility questionnaire. The analysis of misconceptions and the factors that cause them was obtained from the results of the analysis of the five-tier multiple-choice test instrument.

Research Instruments

The data in this study consisted of qualitative and quantitative data. Qualitative data was obtained from interviews with lecturers and students, student needs analysis questionnaires, input from validators regarding the test instruments developed, and instrument feasibility response questionnaires. Quantitative data was used to determine the quality of the instrument developed in terms of validity, feasibility, and effectiveness. Quantitative data in this study were obtained from the results of the assessment of validators, the results of student assessment questionnaires, the results of empirical test instruments used to measure the validity and reliability of instruments, and the results of misconception analysis. All of the instruments have been validated by experts.

Data Analysis

Content validity used the Aiken method with 5 experts consisting of two material experts, two evaluation experts, and one learning practitioner. So, the minimum V value is 0.87 for items declared valid. Aiken's formula is as follows eq (1):

$$V = \frac{S}{[n(c-1)]};\ \text{where } S = \Sigma\, n_i\,(r-\ell_o) \quad (1)$$

Where:
V: Aiken's validity index
ni : number of raters who chose criterion i
c : number of categories/criteria
r : criterion i
ℓo: lowest category
n : total number of raters (Aiken, 1985).

The feasibility of the instrument can be determined by analyzing the results of the student response questionnaire to the instrument as follows eq (2):

$$Ps\ (\%) = \frac{Sum\ of\ scores\ from\ data\ collection}{Total\ score} \times 100\% \quad (2)$$

Minimum feasible if >70%.

Empirical testing used Rasch Model analysis to determine the validity and reliability of the instrument. The validity test was conducted by looking at the MNSQ Outfit value: 0.5<Outfit-MNSQ<1.5; Outfit ZSTD value: -0.2<ZSTD<2; Pt Measure Corr value: 0.4<Point Measure Corr<0.85. If there are items that have MNSQ and PT Measure Corr outfit values outside the accepted values, the ZSTD value must be within the accepted range. The reliability of the question item is determined by looking at the reliability value at the extreme and non-extreme with a minimum score of> 0.67(Ramadhan & Hidayatullah, 2023). Instrument five-tier multiple choice consists of the first level showed a question with five choices and one key answer that students must choose. The second level showed the students' confidence level in choosing answers. The third level showed the reasons students answered the question about the five options provided. The fourth level showed the student's confidence level in selecting a reason. The students' misconceptions were identified based on the students' confidence level in answering diagnostic tests at the second and fourth levels, while the concept of understanding can be identified in the first and third levels. So if the students answer incorrectly in the first and third levels, then they are sure it was included misconception. The fifth level showed sources from student answers (Simamora et al., 2023). Then, the percentage of misconceptions was calculated as follows eq (3):

$$\%\ Misconceptions = \frac{Number\ of\ students\ with\ misconceptions}{total\ students} \times 100\% \quad (3)$$

and the category was determined according to Table 1.

Table 1. Category percentage of misconceptions

Percentage	Category
Low	0% - 30%
Moderate	31% - 60%
High	61% - 100%

RESULT

The development product made in this study is a five-tier multiple-choice test instrument on chemical stoichiometry material. This development step refers to the 4-D model (Define, Design, Develop, and Disseminate) which was developed in this study. (Thiagarajan, S., Semmel & Semmel, 1976). The total number of questions is 15. Define stage, this stage is front-end analysis, learner analysis, task analysis, concept analysis, and analysis of lesson objectives. The design Stage consists of a construction criterion-referenced test, media selection, format selection, and initial design. The development stage consists of expert validation and development testing. The dissemination stage is carried out to promote the product developed. Expert validation results were obtained according to Table 2.

Table 2. Expert Validation results

Number Question	V	Conclusion
1	0,87	Valid
2	0,87	Valid
3	0,93	Valid
4	0,93	Valid
5	0,93	Valid
6	1,00	Valid
7	1,00	Valid
8	1,00	Valid
9	1,00	Valid
10	1,00	Valid
11	1,00	Valid
12	1,00	Valid
13	1,00	Valid
14	1,00	Valid
15	1,00	Valid
16	0,93	Valid
17	1,00	Valid

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.
5	8	46	2.27	.45	1.51	1.9	1.61	1.1	A .21	.48
2	22	46	.16	.36	1.48	2.5	1.42	1.6	B .30	.55
1	22	46	.16	.36	1.38	2.0	1.44	1.6	C .33	.55
10	21	46	.29	.36	1.31	1.6	1.27	1.1	D .39	.56
12	29	46	-.70	.35	1.10	.7	1.10	.4	E .43	.49
11	21	46	.29	.36	.95	-.2	.95	-.1	F .58	.56
4	26	46	-.33	.35	.93	-.4	.89	-.3	G .56	.52
6	21	46	.29	.36	.91	-.5	.85	-.6	H .61	.56
7	21	46	.29	.36	.91	-.5	.88	-.4	I .61	.56
16	11	46	1.72	.41	.90	-.4	.82	-.3	h .58	.53
8	39	46	-2.17	.44	.90	-.3	.60	-.4	g .41	.33
3	23	46	.04	.35	.88	-.7	.79	-.8	f .62	.55
15	30	46	-.83	.36	.87	-.9	.69	-.8	e .57	.48
17	20	46	.42	.36	.85	-.8	.77	-.9	d .65	.56
9	27	46	-.46	.35	.84	-1.1	.70	-1.0	c .62	.51
14	25	46	-.21	.35	.76	-1.6	.73	-1.0	b .67	.53
13	33	46	-1.22	.37	.73	-1.9	.52	-1.1	a .61	.44
MEAN	23.5	46.0	.00	.37	1.01	.0	.94	-.1		
S.D.	7.1	.0	.98	.03	.24	1.3	.31	.9		

Figure 1. Output Misfit Order of Tier-1

Based on the Table 2, all items are declared feasible by material experts, evaluation experts, and educational practitioners. The results analyzed using Rasch to determine the validity (Misfit Order Output) of tier 1 questions can be seen in Figure 1.

The results of the item fit analysis of tier 1 can be seen in Table 3, with inappropriate values in bold and italicized.

Table 3. Results of Tier 1 Item Validity Analysis of the Rasch Model

No	Question item	Measurement accuracy criteria			Decision
		Outfit MNSQ	Outfit ZSTD	PT-MEAN CORR	
1	S5	1,61	1,1	0,21	Valid
2	S2	1,42	1,6	0,30	Valid
3	S1	1,44	1,6	0,33	Valid
4	S10	1,27	1,1	0,39	Valid
5	S12	1,10	0,4	0,43	Valid
6	S11	0,95	-0,1	0,58	Valid
7	S4	0,89	-0,3	0,56	Valid
8	S6	0,85	-0,6	0,61	Valid
9	S7	0,88	-0,4	0,61	Valid
10	S16	0,82	-0,3	0,58	Valid
11	S8	0,60	-0,4	0,41	Valid
12	S3	0,79	-0.8	0,62	Valid
13	S15	0,69	-0,8	0,57	Valid
14	S17	0,77	-0,9	0,65	Valid
15	S9	0,70	-1	0,62	Valid
16	S14	0,73	-1	0,67	Valid
17	S13	0,52	-1	0,61	Valid

The table above shows the results of tier 1, 17 questions, which were all declared valid (fit), so that it can be concluded that the questions met the criteria and guaranteed quality to measure the level of student understanding. The validity of the results (Misfit Order Output) of tier 3 can be seen in Figure 2. The results of the item fit analysis of tier 3 can be seen in Table 4, with the unsuitable values in bold and italicized.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.
17	13	46	1.06	.38	1.57	2.7	2.07	2.1	A .15	.50
13	16	46	.64	.37	1.19	1.1	1.76	2.0	B .37	.53
12	29	46	-1.05	.37	1.45	2.3	1.60	1.7	C .28	.54
9	17	46	.50	.37	1.13	.8	1.32	1.0	D .45	.54
14	22	46	-.14	.36	1.22	1.2	1.26	1.0	E .44	.56
16	16	46	.64	.37	1.23	1.3	1.10	.4	F .42	.53
11	22	46	-.14	.36	1.11	.7	1.09	.4	G .50	.56
7	20	46	.11	.36	1.10	.6	1.02	.2	H .51	.56
3	22	46	-.14	.36	.84	-.9	1.00	.1	I .62	.56
5	27	46	-.78	.36	.93	-.3	.78	-.7	h .61	.55
10	18	46	.37	.36	.91	-.5	.87	-.4	g .59	.55
8	22	46	-.14	.36	.79	-1.2	.88	-.4	f .66	.56
15	8	46	1.88	.44	.81	-.8	.56	-.6	e .53	.42
1	30	46	-1.18	.37	.77	-1.3	.68	-1.0	d .66	.53
4	26	46	-.66	.36	.63	-2.4	.67	-1.3	c .75	.56
6	20	46	.11	.36	.56	-3.0	.46	-2.4	b .80	.56
2	29	46	-1.05	.37	.49	-3.5	.38	-2.5	a .82	.54
MEAN	21.0	46.0	.00	.37	.98	-.2	1.03	.0		
S.D.	5.9	.0	.79	.02	.29	1.7	.45	1.3		

Figure 2. Output Misfit Order Tier-3

Table 4. Results of Tier 3 Item Validity Analysis of the Rasch Model

No	Question item	Measurement accuracy criteria			Decision
		Outfit MNSQ	Outfit ZSTD	PT-MEAN CORR	
1	S17	2,07	2,1	0,15	Invalid
2	S13	1,76	2,0	0,37	Invalid
3	S12	1,60	1,7	0,28	Invalid
4	S9	1,32	1,0	0,45	Valid
5	S14	1,26	1,0	0,44	Valid
6	S16	1,10	0,4	0,42	Valid
7	S11	1,09	0,4	0,50	Valid
8	S7	1,02	0,2	0,51	Valid
9	S3	1,00	0,1	0,62	Valid
10	S5	0,78	-0,7	0,61	Valid
11	S10	0,87	-0,4	0,59	Valid
12	S8	0,88	-0,4	0,66	Valid
13	S15	0,56	-0,6	0,53	Valid
14	S1	0,68	-1,0	0,66	Valid
15	S4	0,67	-1,3	0,75	Valid
16	S6	0,46	-2,4	0,80	Valid
17	S2	0,38	-2,5	0,82	Valid

The table above shows the validity results of tier 3 from 17 questions, 15 questions were declared valid (fit), and 2 questions were declared invalid (misfit). Reliability (summary statistics output) of tier 1 can be seen in Figure 3.

```
     SUMMARY OF 46 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD    MNSQ   ZSTD |
|-------------------------------------------------------------------------------|
| MEAN       8.7      17.0          .12     .63        .99     .1     .94     .1 |
| S.D.       4.2        .0         1.43     .12        .19     .7     .38     .7 |
| MAX.      16.0      17.0         3.20    1.07       1.48    2.4    2.47    2.8 |
| MIN.       2.0      17.0        -2.33     .53        .66   -1.1     .20    -.9 |
|-------------------------------------------------------------------------------|
| REAL RMSE     .66 TRUE SD    1.27  SEPARATION  1.92  Person RELIABILITY  .79 |
|MODEL RMSE     .64 TRUE SD    1.28  SEPARATION  2.00  Person RELIABILITY  .80 |
| S.E. OF Person MEAN = .21                                                     |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .83

     SUMMARY OF 17 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD    MNSQ   ZSTD |
|-------------------------------------------------------------------------------|
| MEAN      23.5      46.0          .00     .37       1.01     .0     .94    -.1 |
| S.D.       7.1        .0          .98     .03        .24    1.3     .31     .9 |
| MAX.      39.0      46.0         2.27     .45       1.51    2.5    1.61    1.6 |
| MIN.       8.0      46.0        -2.17     .35        .73   -1.9     .52   -1.1 |
|-------------------------------------------------------------------------------|
| REAL RMSE     .39 TRUE SD     .90  SEPARATION  2.30  Item   RELIABILITY  .84 |
|MODEL RMSE     .37 TRUE SD     .91  SEPARATION  2.45  Item   RELIABILITY  .86 |
| S.E. OF Item MEAN = .25                                                       |
-------------------------------------------------------------------------------
```

Figure 3. Output Summary Statistic of Tier 1

From the results of the image analysis above, the first table contains a summary of the person's measurements, and the second is a summary of the items seen from the suitability of the items where all items are declared suitable. These results are then summarized and can be seen in Table 5.

Table 5. Reliability Value of Students and Problem Items

Variable	Separation	Reliability	Alpha Cronbach
Person	1,92	0,79	0,83
Question Item	2,30	0,84	

The person reliability value of 0.79 is considered sufficient because it is in the range of 0.67-0.80. The item reliability value of 0.84 is classified as good, in the range of 0.80-0.90, and the alpha Cronbach value is 0.83, meaning that the reliability value between the person and the item is good. The results of the summary statistics output of tier 3 can be seen in Figure 4.

```
     SUMMARY OF 46 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       7.8       17.0       -.27     .63      1.00     .0   1.03     .1 |
| S.D.       4.5         .0       1.46     .12       .14     .7    .34     .7 |
| MAX.      15.0       17.0       2.24    1.04      1.26    1.5   2.16    1.7 |
| MIN.       1.0       17.0      -3.01     .52       .66   -2.3    .48   -2.0 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .66 TRUE SD    1.30  SEPARATION  1.97  Person RELIABILITY  .79 |
|MODEL RMSE    .64 TRUE SD    1.31  SEPARATION  2.04  Person RELIABILITY  .81 |
| S.E. OF Person MEAN = .22                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .85

     SUMMARY OF 17 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE      COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      21.0       46.0        .00     .37       .98    -.2   1.03     .0 |
| S.D.       5.9         .0        .79     .02       .29    1.7    .45    1.3 |
| MAX.      30.0       46.0       1.88     .44      1.57    2.7   2.07    2.1 |
| MIN.       8.0       46.0      -1.18     .36       .49   -3.5    .38   -2.5 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .39 TRUE SD     .69  SEPARATION  1.76  Item   RELIABILITY  .76 |
|MODEL RMSE    .37 TRUE SD     .70  SEPARATION  1.89  Item   RELIABILITY  .78 |
| S.E. OF Item MEAN = .20                                                     |
-------------------------------------------------------------------------------
```

Figure 4. Output Summary Statistic of Tier 3

The results of the figure analysis above are then summarized and can be seen in Table 6.

Table 6. Reliability Value of Students and Problem Items

Variable	Separation	Reliability	Alpha Cronbach
Person	1,97	0,79	0,85
Question Item	1,76	0,76	

The person reliability value of 0.79 is considered sufficient because it is in the range of 0.67-0.80. The item reliability value of 0.76 is considered sufficient because it is in the range of 0.80-0.90, and the Cronbach alpha value is 0.85, meaning that the reliability value between the person and the item is good. The results of the misconception analysis using the five-tier multiple-choice instrument can be seen in Table 7.

Table 7. Misconception Analysis Results

Indicator	2nd Semester	Category	4th Semester	Category
1	53,17	Moderate	38,89	Moderate
2	35,71	Moderate	24,07	Low
3	42,86	Moderate	14,81	Low
4	25,40	Low	28,40	Low
5	33,33	Moderate	18,52	Low
Average	38,09	Moderate	24,94	Low

Misconceptions occurred in all indicators, both among 2nd-semester students and 4th-semester students. The average misconceptions of 2nd semester students are classified as moderate and 4th semester students are classified as low. The causes of student misconception can be seen in Table 8.

Table 8. The Cause of Student Misconception

Aspect	2nd Semester	4th Semester
Personal thinking	52,33	52,60
Internet	34,22	24,94
Book	4,33	1,23
Lecture	9,00	21.23

The biggest factor influencing misconceptions based on the 5TMC test results is personal thinking.

DISCUSSION

The research results show that a five-tier multiple-choice test instrument in stoichiometry material was successfully developed. Define stage, the results of the initial analysis, learner analysis, task analysis, concept analysis, and analysis of lesson objectives found the problem of low student understanding of stoichiometry material. Most students only memorize formulas without understanding the concept. In addition, there is no instrument to detect misconceptions. Students and lecturers feel they need an instrument for misconception analysis, especially in stoichiometry material. Based on this analysis, the concepts in stoichiometry that need to be analyzed for misconceptions are the basic laws of chemistry, the concept of mole, compound composition, empirical and molecular formulas, and reaction equations.

The Design Stage, researchers made a grid of questions and draft questions. The questions were prepared 17 questions consisting of 6 questions of basic laws of chemistry, 4 questions of the mole concept, 1 question of elemental composition, 4 questions of molecular and empirical formulas, and 2 questions of reaction equations. The number of questions for each concept depends on the number of misconceptions found in the previous literatures. This instrument was designed by Google Forms.

The development stage consists of expert validation and development testing. All items are declared valid by material experts, evaluation experts, and educational practitioners. Then, the draft question also was revised according to expert advice and became draft question 2. Draft question 2 was used for small-scale trials tested for 9 chemistry students in the sixth semester to determine the feasibility of the question. The results of the response questionnaire amounted to 83.12% so the instrument was declared feasible. The next step in this stage is development testing. This instrument was tested on 2nd-semester biology students and analyzed using Rasch to determine validity and reliability. All questions of tier 1 were declared valid, and tier 3 had 15 questions declared valid, and 2 questions were declared invalid (question numbers 13 and 17). Questions number 17 and 13 did not match according to the criteria for Outfit MNSQ, Outfit ZSTD, or PT-MEAN CORR. So, it can be concluded that the 15 questions that were declared valid met the criteria and guaranteed quality to be used to measure the level of student understanding. Reliability is the interaction between a person and an item, which can be seen from the Cronbach alpha value. The minimum reliability value is 0.67. Reliability results of tier 1 showed that the person reliability value of 0.79 is considered sufficient because it is in the range of 0.67-0.80. The item reliability value of 0.84 is classified as good, in the range of 0.80-0.90, and the alpha Cronbach value is 0.83, meaning that the reliability value between the person and the item is good. Tier 3 showed that the person reliability value of 0.79 is considered sufficient because it is in the range of 0.67-0.80. The item reliability value of 0.76 is considered sufficient because it is in the range of 0.80-0.90, and the Cronbach alpha value is 0.85, meaning that the reliability value between the person and the item is good. Based on the validity and reliability analysis results, the valid and reliable question items totaled 15 questions, which will be used for the large-scale trial. Question items that are not used are numbers 13 and 17 with consideration of the validity of the question. A large-scale trial was conducted to obtain evidence of the effectiveness of the final product in analyzing misconceptions. This trial was conducted on chemistry students in the 2nd and 4th semesters of the 2023/2024 academic year. Misconceptions occurred in all indicators, both among 2nd-semester students and 4th-semester students. The average misconceptions of 2nd semester students are classified as moderate and 4th semester students are classified as low. The highest misconceptions occur in the concept of basic laws of chemistry. Most students know and memorize the sound of all the basic laws of chemistry. However, most of them do not understand the concept of the basic law. This is in line with research conducted by (Lahinda & Krisen, 2021; Kristyowati & Priatmoko, 2023; Rina Elvia et al., 2022). The biggest factor influencing misconceptions based on the 5TMC test results is personal thinking. This aligns with research conducted by (Resbiantoro et al., 2022; Nufus & Silfianah, 2023). The cause of misconceptions that come from personal thinking occur because students misinterpret their understanding or thoughts about a concept.

The dissemination stage is carried out to promote the product developed. The lecturers and students were very interested in watching the research presentations. This can be seen from the many questioners, and they want to make a five-tier instrument to analyze misconceptions in other chemical materials. Even students are also very happy because they know the misconceptions they have experienced in stoichiometry material. In this stage, lecturers and students also have a discussion about how to remediate stoichiometry to reduce misconceptions.

CONCLUSION

The Five-tier multiple choice test instrument developed was declared feasible according to content validity, Rasch Model, and response questionnaire. This instrument is also effective for analyzing misconceptions in chemistry students in semester 2 with a moderate category and semester 4 with a low category. In addition, this instrument can also be used to determine the factors that influence misconceptions. The biggest factor influencing misconceptions based on the 5TMC test results is personal thinking. The Five-tier multiple choice test instrument can be an alternative to analyze misconceptions and sources of misconception in stoichiometry material. It can be developed for other materials. It is hoped that remediation steps can be taken to eliminate these misconceptions. From this research, five-tier multiple choice only analyzed student understanding from aspects of ontology and epistemology, and it is suggested that further research is needed to develop a six-tier multiple-choice instrument to analyze students' understanding comprehensively from aspects of ontology, epistemology, and axiology.

ACKNOWLEDGMENTS

Thank you to Siber Syekh Nurjati Islamic University for supporting financially to this research and all of the people who helped me to do this research.

REFERENCES

Aiken, L. (1985). Three coefficient for analyzing the realibility and validity of rating. *Educational and Psychological Measurement*, *45*, 131–132.

Barke H D, H. A. (2009). *Misconceptions in chemistry: Addressing perceptions*. Heidelberg: Springer - Verlag.

Drastisianti, A., Susilaningsih, E., Supartono, M., & Wijayati, N. (2018). *The Study of Chemistry Learning on The Material of Buffer Solution Supported by Teaching Material of Multiple Representation-Chemoentrepreneurship Viewed From Student Entrepreneurship Interest*. *247*(Iset), 27–31. https://doi.org/10.2991/iset-18.2018.6

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Härmälä-Braskén, A. S., Hemmi, K., & Kurtén, B. (2020). Misconceptions in chemistry among Finnish prospective primary school teachers–a long-term study. *International Journal of Science Education*, *42*(9), 1447–1464. https://doi.org/10.1080/09500693.2020.1765046

Jammeh, A. L. J., Karegeya, C., & Ladage, S. (2023). Misconceptions on Basic Stoichiometry Among the Selected Eleventh-Grade Students in the Urban Regions of the Gambia. *Journal of Baltic Science Education*, *22*(2), 254–268. https://doi.org/10.33225/jbse/23.22.254

Johnstone, A. H. (2000). Teaching of Chemistry - Logical or Psychological? *Chem. Educ. Res. Pract.*, *1*(1), 9–15. https://doi.org/10.1039/a9rp90001b

Kanwal, S., & Farooq, M. S. (2021). Development of Diagnostic Assessment Test for Misconceptions (ATM) in Mathematics at Elementary School Level. *Global Educational Studies Review*, *VI*(I), 94–102. https://doi.org/10.31703/gesr.2021(vi-i).10

Kristyowati, V., & Priatmoko, S. (2023). Analisis Miskonsepsi Stoikiometri Melalui Tes Diagnostik Berbantuan Certainty of Response Index. *Chemistry in Education*, *12*(1), 67–75. http://journal.unnes.ac.id/sju/index.php/chemined

Lahinda, C. G., & Krisen, S. S. (2021). Analisis Miskonsepsi Siswa Pada Materi Hukum Dasar Kimia Menggunakan Two-Tier Diagnostic Test Di SMA Negeri 1 Tatapaan. *General Chemistry Journal*, *1*(1), 5–10. https://doi.org/10.37033/ojce.v3i1.268

Lestari, E. A., Harjito, H., Susilaningsih, E., & Wijayati, N. (2021). Analisis Miskonsepsi Menggunakan Tes

Diagnosa Three-Tier Multiple Choice Pada Materi Stoikiometri. *Jurnal Inovasi Pendidikan Kimia*, *15*(2), 2824–2830. https://doi.org/10.15294/jipk.v15i2.29171

Malaterre, C., Javaux, E. J., & López-García, P. (2023). Misconceptions in Science. *Perspectives on Science*, *31*(6), 717–743. https://doi.org/10.1162/posc_a_00590

Mamombe, C., Mathabathe, K. C., & Gaigher, E. (2021). Teachers' and Learners' Perceptions of Stoichiometry using POGIL: A Case Study in South Africa. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(9), 1–11. https://doi.org/10.29333/ejmste/11140

Mardeni, P. R. (2023). Review Artikel : Five-Tier Multiple Choice untuk Mendeteksi Penyebab Miskonsepsi Kesetimbangan Kimia Review Article : Five-Tier Multiple Choice to Detect the Causes of Chemical Equilibrium Misconceptions. *Seminar Nasional Kimia (SNK) FMIPA Universitas Negeri Surabaya*, *September*, 50–61.

Mufit, F., Festiyed, F., Fauzan, A., & Lufri, L. (2018). Impact of Learning Model Based on Cognitive Conflict toward Student's Conceptual Understanding. *IOP Conference Series: Materials Science and Engineering*, *335*(1). https://doi.org/10.1088/1757-899X/335/1/012072

Nufus, S., & Silfianah, I. (2023). Analisis Miskonsepsi Peserta Didik pada Materi Struktur Atom Menggunakan Five-Tier Multiple Choice Diagnostic Test Berbasis Piktorial. *Orbital: Jurnal Pendidikan Kimia*, *7*(2), 126–139. https://doi.org/10.19109/ojpk.v7i2.19239

Opara, M. F. (2014). Improving Students Performance in Stoichiometry through the Implementation of Collaborative Learning. *Journal of Education and Vocational Research*, *5*(3), 85–93. https://doi.org/10.22610/jevr.v5i3.156

Polamolo, A. N., & Lukum, A. (2022). Analisis Miskonsepsi Materi Pereaksi Pembatas Menggunakan Tes Three Tier Multiple Choice. *Jambura Journal of Educational Chemistry*, *4*(1), 51–56. https://doi.org/10.34312/jjec.v4i1.13180

Rahayu, I., . W., Widhiyanti, T., & Mulyani, S. (2024). Analysis of Misconceptions on the Factors that Affect the Reaction Rate. *International Conference On Mathematics And Science Education*, *2024*, 140–150. https://doi.org/10.18502/kss.v9i13.15914

Ramadhan, A. F., & Hidayatullah, R. S. (2023). Analisis Kualitas Butir Soal Ujian Satuan Pendidikan (Usp) Materi C2 Teknik Pemesinan Kelas Xii Di Smk Pgri 1 Lamongan Melalui Model Rasch. *Jurnal Pendidikan Teknik Mesin*, *12*(3), 1–10.

Resbiantoro, G., Setiani, R., & Dwikoranto. (2022). A Review of Misconception in Physics: The Diagnosis, Causes, and Remediation. *Journal of Turkish Science Education*, *19*(2), 403–427. https://doi.org/10.36681/tused.2022.128

Rina Elvia, Amelia, T., & Handayani, D. (2022). Identifikasi Miskonsepsi Siswa Pada Pembelajaran Kimia Menggunakan Metode Four-Tier Diagnostik Test Di Sma Negeri 03 Kota Bengkulu. *Alotrop, Jurnal Pendidikan Dan Ilmu Kimia*, *6*(2), 110–117. https://doi.org/10.33369/alo.v6i2.25099

Sa'adah, N., Haryani, S., & Mahatmanti, W. (2022). Analisys of Chemical Misconceptions of Stoichiometry Materials Using Certainty of Response Index. *International Journal of Active Learning*, *7*(1), 94–102.

Salame, I. I., Casino, P., & Hodges, N. (2020). Examining Challenges that Students Face in Learning Organic Chemistry Synthesis. *International Journal of Chemistry Education Research*, *4*(April), 1–9. https://doi.org/10.20885/ijcer.vol4.iss1.art1

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

Shiddiqi, M. H. A., Arthamena, V. D., Ayyubi, M., Manarisip, A. J., & Aznam, N. (2024). Systematic Literature Review: Analysis of Misconception Problems and Diagnostic Instruments for Learning Chemistry. *Jurnal Penelitian Pendidikan IPA*, *10*(4), 168–179. https://doi.org/10.29303/jppipa.v10i4.5189

Simamora, R., Maison, & Kurniawan, W. (2023). Identifikasi Miskonsepsi Peserta Didik Menggunakan Five-Tier Diagnostic Test pada Materi Fluida Statis di SMAN 7 Kota Jambi. *Jurnal Penelitian Pendidikan Fisika*, *8*(2), 139–144. https://doi.org/10.36709/jipfi.v8i2.18

Siswaningsih, W., Nahadi, & Widasmara, R. (2019). Development of Three Tier Multiple Choice Diagnostic Test to Assess Students' Misconception of Chemical Equilibrium. *Journal of Physics: Conference*

Series, *1280*(3), 5–10. https://doi.org/10.1088/1742-6596/1280/3/032019
Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649
Soeharto, S., & Csapó, B. (2022). Exploring Indonesian student misconceptions in science concepts. *Heliyon*, *8*(9). https://doi.org/10.1016/j.heliyon.2022.e10720
Supatmi, S., Setiawan, A., Rahmawati, Y., Education, C., Program, S., & Jakarta, U. N. (2019). Students' misconceptions of acid-base titration assessments using a two - tier multiple-choice diagnostic test. *African Journal of Chemical Education*, *9*(1), 18–37.
Taha, H., Hashim, R., Jusoff, K., & Khoo, Y. Y. (2014). The Influence of Students' Concept of Mole, Problem Representation Ability and Mathematical Ability on Stoichiometry Problem Solving Islamic Integrated Curriculum View project Philosophy for children View project. *WEI International Academic Conference Proceedings*, 122–136. http://scottishjournal.co.uk
Thiagarajan, S., Semmel, D. S. dan, & Semmel, M. I. (1976). *Instructional development for training teachers of exceptional children: A sourcebook*. University of Minnesota. https://doi.org/10.1016/0022-4405(76)90066-2
Üce, M., & Ceyhan, İ. (2019). Misconception in Chemistry Education and Practices to Eliminate Them: Literature Analysis. *Journal of Education and Training Studies*, *7*(3), 202. https://doi.org/10.11114/jets.v7i3.3990
Woldeamanuel,M., Atagana,H., & Engida, T. (2014). What Makes Chemistry Difficult. *African Journal of Chemistry Education (AJCE)*, *4*(2), 31–43.