Assimilation: Indonesian Journal of Biology Education
ISSN 2621-7260 (Online)
Journal homepage: https://ejournal.upi.edu/index.php/asimilasi

Development of higher order thinking skill questions using Stahl and Murphy's taxonomy on excretion system topic

Cahya Gita Camila[1], Afandi[1*], Andi Besse Tenriawaru[1], Wiwit Artika[2], Nurfadilah Siregar[3]
[1] Biology Education Study Program, FKIP, Universitas Tanjungpura, Prof. Dr. H. Hadari Nawawi Street, Pontianak
[2] Biology Education Study Program, FKIP, Universitas Syiah Kuala, Teuku Nyak Arief Street, Darussalam, Banda Aceh
[3] Mathematics Education Study Program, FKIP, Universitas Tanjungpura, Prof. Dr. H. Hadari Nawawi Street, Pontianak
*Corresponding author: afandi@fkip.untan.ac.id

OPEN ACCESS

ARTICLE HISTORY
Received: 20 July 2023
First Revised: 13 September 2023
Accepted: 30 September 2023
First Available Online: 30 September 2023
Publication Date: 30 September 2023

KEYWORDS
Higher order thinking skills
Excretory system
Stahl and Murphy's taxonomy

ABSTRACT

This study aims to determine the development of higher-order thinking skills questions using Stahl's and Murphy's Taxonomy on excretory system material for 8th grade of junior high school. The subjects of this study were 305 8th grade students of Junior High School 2 Sintang. This study uses the Research and Development method. The sampling technique uses total sampling. The instrument used is a multiple-choice test consisting of 50 questions on excretory system material. The results of this study based on the validity of items showed that the number of questions accepted was 50 items. Reliability analysis showed Cronbach's alpha value of 0.71, person reliability value of 0.71, and item reliability of 0.94. Through this research, we found that the instrument of higher-order thinking skills using Stahl and Murphy's taxonomy on excretion system material for 8th grade of junior high school can be said to be valid and reliable.

INTRODUCTION

The application of 21st Century learning expects a teacher to have the ability and understanding of TPACK (Pedagogical Content Knowledge) (Shafie et al., 2019; Zhou et al., 2023) where this ability will be more makes it easier to hone 4C skills and master and understand the technology that is a demand in the 21st Century (Darry et al, 2021). Critical thinking, communication, collaboration and creativity (4C) are considered in accordance with the demands of today's era and teaching staff are also expected to be able to understand the Higher Order Thinking Skills (HOTS) in 21st Century learning to make it easier to hone students' evaluation and creative abilities (Ichsan et al., 2019; Wening & Santosa, 2020). The government hopes that students can achieve various competencies by applying higher order thinking skills or HOTS. These competencies include critical thinking, creative and innovative, the ability to work together and self-confidence. These five things are the targets of student character in the National Examination evaluation system and are also skills in the 21st Century (Mahanal et al., 2019).

The problems experienced by students in facing HOTS questions in science subjects are very diverse. The problem of solving HOTS questions has also been raised by the International Program for International Student Assessment (PISA) which states that the achievement of reading literacy, mathematics literacy and scientific literacy achieved by students in Indonesia is very low (Saputri et al., 2019; Suwono et al., 2023). The low achievement of these students, especially in solving contextual problems, requires reasoning, argumentation and creativity in solving them. The questions included in the Trends in International Mathematics and Science Study (TIMSS) are HOTS-based questions (Fanani, 2018; Markandan & Osman, 2023).

One of the things that encourages the creation of superior graduates is the learning process (Sayekti, 2019). The Indonesian Ministry of Education is currently making changes to the 2013 Curriculum, namely the 2013 Curriculum, revised 2017. This curriculum aims at developing 21st Century skills and high-level thinking skills (Rosdiana & Pahlevi, 2020; Widiawati et al., 2018). Thinking at a higher level of cognition is an indispensable skill in the learning process. Therefore, to increase the International Program for International Student Assessment (PISA) score and create quality Indonesian human resources, improvements and updates are needed in the educational aspect (Susilowati et al., 2022).

Higher order thinking skills (HOTS) are thinking skills that are more than just memorizing facts or concepts. Higher order thinking skills require students to do something about these facts. Students must understand, analyze each other, categorize, manipulate, create new ways creatively, and apply them in finding solutions to new problems (Jaenudin et al., 2020). The low level of high-level thinking skills (HOTS) was stated by Wikanta et al. (2022) who stated that the problems that occur in schools are that the questions or questions used in cognitive assessments tend to test more on the memory aspect, while for questions that train higher order thinking skills (HOTS) tend to be absent. A teacher in training students to work on questions characterized by high level thinking (HOTS) can develop an assessment instrument based on high level thinking (HOTS). Added by Susantini et al. (2022) through the development of assessment instruments based on higher order thinking (HOTS), it can advance the quality of human resources in Indonesia.

Based on the results of research conducted by Sari et al. (2021) which states that the process of assessing student learning outcomes requires instruments that must be prepared and paid attention to first so that learning objectives can be achieved optimally. The application of developing high-level thinking test instruments (HOTS) is able to develop students' high-level thinking skills, as well as providing examples of high-level thinking (HOTS) questions to teachers. This problem can be overcome by developing test instruments based on higher order thinking (HOTS). The higher order thinking test instrument (HOTS) developed is a valid and reliable test instrument. In line with previous research conducted by Yusuf et al. (2021) stated that the assessment instrument based on higher order thinking (HOTS) for high school students in

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

Surakarta had validity and reliability values with high criteria, then the class that applied the HOTS-based assessment instrument in the learning process can measure students' high-level thinking.

The level of learning evaluation test that is often used is Bloom's taxonomy. Bloom's taxonomy is a multilevel structure that identifies thinking skills from low to high. The low level of high-level thinking skills is in line with research conducted by Irawati (2018) using Bloom's taxonomy showing that the maximum score is 100%, students' analytical skills reach 30%, their evaluating level reaches 32%, and their creation level reaches 23%. Based on this percentage value, it shows that the ability to solve high level thinking (HOTS) questions is still low. However, research using Stahl and Murphy's taxonomy is rare. Based on the definition of Stahl and Murphy's taxonomy, it can measure more complex higher order thinking skills, that is, it can be used in international learning planning and it can be assumed that teachers can conclude from students' behavior as a mental process during learning.

Stahl and Murphy's taxonomy system is an ambitious attempt to create a framework for classifying learning objectives, taking learning patterns and linear thinking from information to creative synthesis, from ideas and beliefs (Koman et al., 2023). The cognition domain describes the sequence of degrees or levels of information and rules that can be internalized and used by students. Therefore, several levels of thinking and learning are included in the HOTS level in Stahl and Murphy's taxonomy, namely transformation (information transformation), transfusion (information transfer), incorporation (information merging), organization (information organization), and generation (generating information) (Afandi & Ningsih, 2020; Zain et al., 2022).

To analyze question items using Rasch modeling, Winsteps software can be used. Winsteps software is a computational tool using the Rasch model to analyze scores produced from test instruments with the aim of knowing MNSQ Outfit, ZSTD Outfit, Point Measure Correlation, Item Reliability and Cronbach's Alpha. MNSQ Outfit is useful for seeing the suitability of the data to the model used. The expected Mean Square value is 1 (one). If the Mean-Square value in infit is greater than 1, the variation of the instrument is greater than the prediction made by the Rasch model. If the infit value is less than 1, then there will be fewer variations in the instrument carried out by the Rasch model (Azizah & Wahyuningsih, 2020).

In the 21st century, high-level thinking skills are considered important, so questions need to be developed that can improve students' high-level thinking skills. For this reason, researchers will develop questions that measure high-level thinking skills. The questions in this study consisted of 50 multiple choice questions regarding the excretory system material. The excretory system material is part of the science material taught to students in 8th grade of even semester middle school, which is material that is quite difficult to understand in science learning because it consists of quite a lot of concepts and example questions (Riyani & Siregar, 2022). Therefore, researchers will develop high-level thinking skills questions using Stahl and Murphy's taxonomy in excretory system topic for 8th grade of junior high school.

METHODS

The method used in this research is the Borg & Gall's Research & Development (R&D) method, include potential and problems, data collection, product design, design validation, design revision, and product testing (Setiawan et al., 2021). The instrument developed in this research is a high-level thinking skills test instrument consisting of 50 multiple choice questions consisting of one answer key and three distractors. The questions developed regarding excretory system material refer to Stahl and Murphy's Taxonomy which consists of 5 levels of HOTS, namely transformation (information transformation), transfusion (information transfer), incorporation (information merging), organization (information organization), and generation (generating information).

Indicators at the transformation level (information transformation) are proposing, interpreting, classifying, summarizing and evaluating. Indicators of transfusion (transfer of

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

information) are estimating, confirming, differentiating, abstracting and converting. Indicators for incorporation (combining information) are explaining, selecting, organizing, solving and interpreting. Indicators in organization (information organization) are starting, carrying out, modifying, completing and assessing. Indicators for generation (generating information) are designing, compiling, creating, producing and developing. This research was conducted on 8th grade students at Junior High School 2 Sintang. The selection of respondents in this study used a total sampling technique totaling 305 students from 8th grade of class A to J. After respondents filled in the high-level thinking skills questions given, then item analysis was carried out from the instrument using the Rasch model with the Winsteps' software tools.

RESULTS AND DISCUSSION

Item validity

Item validity was carried out on 50 questions with the aim of knowing which items were accepted and which items were rejected. Item validity can be obtained by testing directly on respondents. The validity of items for the higher order thinking skills test instrument was analyzed using the Winsteps program with Rasch Model. The Rasch model is a development of an analytical model by George Rasch from response theory item 1 PL (one Logistic Parameter). The steps that can be taken to display the item fit table are to click output tables, then click table 10 item fit order on the main menu. The output from the item fit results on the instrument used can be seen in the analysis results (Figure 1).

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
39	73	305	.79	.14	1.08	1.1	1.18	1.9	A .09	.24	76.1	76.5	S39
28	110	305	.17	.12	1.09	2.1	1.14	2.5	B .10	.26	61.3	66.0	S28
48	74	305	.77	.14	1.06	.8	1.13	1.5	C .12	.24	77.0	76.2	S48
40	82	305	.62	.13	1.03	.5	1.10	1.3	D .17	.24	74.1	73.8	S40
47	110	305	.17	.12	1.06	1.3	1.08	1.5	E .16	.26	65.2	66.0	S47
14	87	305	.54	.13	1.02	.4	1.07	1.0	F .19	.25	74.4	72.3	S14
42	80	305	.66	.13	1.07	1.1	1.07	.9	G .12	.24	72.1	74.4	S42
41	116	305	.08	.12	1.04	1.0	1.07	1.4	H .19	.26	62.0	64.6	S41
30	112	305	.14	.12	1.05	1.2	1.06	1.2	I .17	.26	62.0	65.5	S30
17	122	305	-.01	.12	1.03	.7	1.06	1.2	J .21	.26	61.3	63.3	S17
37	105	305	.24	.12	1.05	1.1	1.03	.6	K .18	.25	62.3	67.2	S37
50	120	305	.02	.12	1.02	.4	1.05	1.1	L .22	.26	64.6	63.7	S50
34	129	305	-.11	.12	1.05	1.4	1.05	1.2	M .18	.26	60.3	62.1	S34
44	130	305	-.13	.12	1.04	1.3	1.04	1.0	N .19	.26	58.0	62.0	S44
23	81	305	.64	.13	1.01	.2	1.04	.6	O .22	.24	75.1	74.1	S23
24	109	305	.18	.12	1.03	.8	1.04	.8	P .20	.26	68.2	66.2	S24
31	104	305	.26	.12	1.03	.6	1.04	.7	Q .20	.25	65.2	67.5	S31
43	121	305	.00	.12	1.04	1.0	1.04	.8	R .20	.26	62.3	63.5	S43
38	81	305	.64	.13	1.03	.6	1.03	.4	S .19	.24	73.1	74.1	S38
25	134	305	-.19	.12	1.02	.8	1.03	.7	T .22	.26	61.6	61.4	S25
35	96	305	.39	.13	1.02	.5	1.02	.3	U .21	.25	68.9	69.7	S35
45	96	305	.39	.13	1.01	.3	1.02	.4	V .22	.25	70.8	69.7	S45
15	105	305	.24	.12	1.01	.2	1.01	.3	W .24	.25	67.5	67.2	S15
49	146	305	-.36	.12	1.01	.5	1.01	.2	X .24	.26	59.3	60.6	S49
22	136	305	-.21	.12	1.01	.4	1.01	.2	Y .25	.26	59.7	61.2	S22
27	118	305	.05	.12	1.01	.3	1.00	.0	y .25	.26	63.3	64.1	S27
33	103	305	.27	.13	.99	-.2	.99	-.1	x .27	.25	66.6	67.8	S33
19	107	305	.21	.12	.99	-.2	.98	-.3	w .27	.26	67.5	66.7	S19
29	84	305	.59	.13	.99	-.1	.98	-.2	v .26	.24	74.1	73.2	S29
26	169	305	-.68	.12	.98	-.6	.99	-.3	u .29	.26	63.0	61.5	S26
1	195	305	-1.06	.12	.99	-.3	.98	-.4	t .28	.25	66.9	66.0	S1
6	107	305	.21	.12	.98	-.5	.98	-.3	s .29	.26	68.2	66.7	S6
11	97	305	.37	.13	.98	-.3	.97	-.5	r .28	.25	69.2	69.4	S11
7	125	305	-.06	.12	.97	-1.0	.98	-.4	q .31	.26	66.2	62.8	S7
9	135	305	-.20	.12	.97	-1.0	.98	-.5	p .31	.26	65.9	61.3	S9
36	90	305	.49	.13	.98	-.3	.94	-.8	o .29	.25	70.2	71.4	S36
32	94	305	.42	.13	.98	-.4	.96	-.6	n .29	.25	72.8	70.3	S32
13	133	305	-.17	.12	.97	-.8	.96	-1.0	m .31	.26	63.3	61.6	S13
16	185	305	-.91	.12	.97	-.7	.96	-.9	l .31	.26	66.6	63.9	S16
3	146	305	-.36	.12	.96	-1.3	.96	-1.0	k .32	.26	63.3	60.6	S3
18	201	305	-1.15	.12	.95	-1.2	.96	-.6	j .33	.25	69.2	67.5	S18
4	216	305	-1.40	.13	.96	-.8	.92	-1.0	i .32	.24	73.1	71.6	S4
20	104	305	.26	.12	.95	-1.2	.96	-.8	h .34	.25	71.1	67.5	S20
21	117	305	.06	.12	.95	-1.4	.94	-1.3	g .35	.26	66.9	64.4	S21
46	140	305	-.27	.12	.95	-1.8	.93	-1.8	f .36	.26	63.6	60.9	S46
12	157	305	-.51	.12	.94	-2.2	.92	-2.1	e .37	.26	63.6	60.6	S12
2	110	305	.17	.12	.93	-1.7	.92	-1.5	d .37	.26	68.5	66.0	S2
5	129	305	-.11	.12	.92	-2.4	.92	-2.0	c .39	.26	67.5	62.1	S5
8	222	305	-1.50	.13	.91	-1.4	.87	-1.7	b .39	.24	75.7	73.3	S8
10	165	305	-.62	.12	.90	-3.4	.88	-3.3	a .43	.26	66.9	61.1	S10
MEAN	122.2	305.0	.00	.12	1.00	-.1	1.01	.0			67.1	66.6	
S.D.	35.4	.0	.53	.01	.04	1.1	.06	1.1			4.9	4.6	

Figure 1. Item fit output in Winsteps

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

The level of item fit (Item Fit) can be seen using several criteria, namely the outfit mean-square value, outfit z-standard, and point measure correlation (Nugroho, 2022; Purnami et al., 2021). Rasch Item Fit modeling is used to see whether an instrument item can measure normally. Point Measure Correlation (Pt Measure Corr) value: 0.4 < Point Measure Corr < 0.85. Outfit mean square (MNSQ) value accepted: 0.5 < MNSQ < 1.5. Outfit Z standard (ZSTD) value accepted: –2.0 < ZSTD < +2.0. If the question items in these three criteria are not met, it can be ascertained that the question items are not appropriate and therefore need to be updated. At a minimum, a question item is said to be fit if it meets 1 of the 3 criteria. The results of item validity can be seen in (Table 1).

Based on (Table 1), it is known that of the 50 questions, 50 questions were also accepted (S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, S36, S37, S38, S39, S40, S41, S42, S43, S44, S45, S46, S47, S48, S49, S50). If it is converted into a percentage, 100% of the questions are accepted or fit.

Based on the 50 questions created, there were misfit items (items that had fit statistical values that were too high or too low) in the instrument being analyzed. The Pt. Measure Coor in this study for all question items has met the criteria. However, there are several questions that do not fit into one of the criteria for Outfit MNSQ, namely questions number 39, 48, 40, 47, 14, 42, 41, 30, and 17. Next in Outfit ZSTD are questions number 10 and 12. Questions that are not included in the two criteria in the ZSTD Outift is question number 28.

Uddin (2021) and Wihardjo (2021) states that the value of Outfit mean-square, Outfit z-standard, and Point Measure Correlation are the criteria used to see the level of suitability of items (item fit). Outfit mean square (MNSQ) value accepted: 0.5 < MNSQ < 1.5. The value for the Outfit Z-standard (ZSTD) received: -2.0 < ZSTD < +2.0 and the Point Measure Correlation (Pt Measure Corr) value: 0.4 < Point Measure Corr < 0.85. If the question items in these three criteria are not met, then it can be said that the question items are not good enough and need to be repaired or replaced. However, if the item meets one of the three criteria, the item can be said to be fit, and does not need to be repaired or replaced.

Table 1. Item validity results

Item No.	Outfits		Pt. Measure Corr	Interpretation
	MNSQ	ZSTD		
1	0.98	-0.4	0.28	Accepted
2	0.92	-1.5	0.37	Accepted
3	0.96	-1.0	0.32	Accepted
4	0.92	-1.0	0.32	Accepted
5	0.92	-2.0	0.39	Accepted
6	0.98	-0.3	0.29	Accepted
7	0.98	-0.4	0.31	Accepted
8	0.87	-1.7	0.39	Accepted
9	0.98	-0.5	0.31	Accepted
10	0.88	-3.3	0.43	Accepted
11	0.97	-0.5	0.28	Accepted
12	0.92	-2.1	0.37	Accepted
13	0.96	-1.0	0.31	Accepted
14	1.07	1.0	0.19	Accepted
15	1.01	0.3	0.24	Accepted
16	0.96	-0.9	0.31	Accepted
17	1.06	1.2	0.21	Accepted
18	0.96	-0.6	0.33	Accepted
19	0.98	-0.3	0.27	Accepted
20	0.96	-0.8	0.34	Accepted

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

Item No.	Outfits		Pt. Measure Corr	Interpretation
	MNSQ	ZSTD		
21	1.01	0.2	0.14	Accepted
22	0.94	-1.3	0.35	Accepted
23	1.04	0.6	0.22	Accepted
24	1.04	0.8	0.20	Accepted
25	1.03	0.7	0.22	Accepted
26	0.99	-0.3	0.29	Accepted
27	1.00	0.0	0.25	Accepted
28	1.14	2.5	0.10	Accepted
29	0.98	-0.2	0.26	Accepted
30	1.06	1.2	0.17	Accepted
31	1.04	0.7	0.20	Accepted
32	0.96	-0.6	0.29	Accepted
33	0.99	-0.1	0.27	Accepted
34	1.05	1.2	0.18	Accepted
35	1.02	0.3	0.21	Accepted
36	0.94	-0.8	0.29	Accepted
37	1.03	0.6	0.18	Accepted
38	1.03	0.4	0.19	Accepted
39	1.18	1.9	0.09	Accepted
40	1.10	1.3	0.17	Accepted
41	1.07	1.4	0.19	Accepted
42	1.07	0.9	0.12	Accepted
43	1.04	0.8	0.20	Accepted
44	1.04	1.0	0.19	Accepted
45	1.02	0.4	0.22	Accepted
46	0.93	-1.8	0.36	Accepted
47	1.08	1.5	0.16	Accepted
48	1.13	1.5	0.12	Accepted
49	1.01	0.2	0.24	Accepted
50	1.05	1.1	0.22	Accepted

Difficulty level of question items (item measure), and individual ability level

The results of the analysis of the level of difficulty of the questions in the research were carried out using the Winsteps program and the results were obtained to find out which questions were categorized as very difficult, difficult, easy and very easy. The categories of measure values can be seen in (Table 2) and the results of the difficulty level of the items that have been analyzed can be seen in (Table 3).

Based on (Table 2) above, it is known that 6 questions (S23, S38, S39, S40, S42, and S48) are categorized as very difficult questions with a percentage of 12%. There are 24 questions (S2, S6, S11, S14, S15, S19, S20, S21, S24, S27 S28, S29, S30, S31, S32, S33, S35, S36, S37, S41, S43, S45, S47, and S50) which is included in the difficult question category with a percentage of 48%. 13 questions from the validated items (S3, S5, S7, S9, S12, S13, S17, S22, S25, S34, S44, S46, and S49) are included in the easy question category with a percentage of 26%. 7 questions out of 50 validated questions (S1, S4, S8, S10, S16, S18, and S26) fall into the very easy question category with a percentage of 14%.

Individual abilities are divided into three categories, namely high, medium and low. Categorization was obtained based on: person measure > SD: high, SD < person measure < mean: medium, and < mean: low. The SD value obtained in this analysis was 0.59 and the mean value was -0.45 (Fitriani et al., 2019; Siburian et al., 2019).

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

Table 2. Measure value categories

Score Category	Degree of difficulty
Smaller than -1SD or smaller than the -SD value	Very easy
0.0 logit – 1 SD or -SD < 0.0	Easy
0.0 logit + 1SD or 0.0 to SD value	Difficult
> +1SD or > SD value	Very difficult

Information: SD = Standard Deviation

Table 3. Results of difficulty level of question items

HOTS level	Indicator	No. Question	Measure Logit	Interpretation
Transformation	Propose	1	-1.06	Very easy
Transformation	Propose	2	0.17	Difficult
Transformation	Interpret	3	-0.36	Easy
Transformation	Interpret	4	-1.40	Very easy
Transformation	Classifying	5	-0.11	Easy
Transformation	Classifying	6	0.21	Difficult
Transformation	Summarizing	7	-0.06	Easy
Transformation	Summarizing	8	-1.50	Very easy
Transformation	Evaluate	9	-0.20	Easy
Transformation	Evaluate	10	-0.62	Very easy
Transfusion	Estimate	11	0.37	Difficult
Transfusion	Estimate	12	-0.51	Easy
Transfusion	Confirm	13	-0.17	Easy
Transfusion	Confirm	14	0.54	Difficult
Transfusion	Differentiate	15	0.24	Difficult
Transfusion	Differentiate	16	-0.91	Very easy
Transfusion	Abstracting	17	-0.01	Easy
Transfusion	Abstracting	18	-1.15	Very easy
Transfusion	Converting	19	0.21	Difficult
Transfusion	Converting	20	0.26	Difficult
Incorporation	Explain	21	0.06	Difficult
Incorporation	Explain	22	-0.21	Easy
Incorporation	Choose	23	0.64	Very difficult
Incorporation	Choose	24	0.18	Difficult
Incorporation	Arrange	25	-0.19	Easy
Incorporation	Arrange	26	-0.68	Very easy
Incorporation	Solve	27	0.05	Difficult
Incorporation	Solve	28	0.17	Difficult
Incorporation	Interpret	29	0.59	Difficult
Incorporation	Interpret	30	0.14	Difficult
Organization	Start	31	0.26	Difficult
Organization	Start	32	0.42	Difficult
Organization	Do	33	0.27	Difficult
Organization	Do	34	-0.11	Easy
Organization	Modify	35	0.39	Difficult
Organization	Modify	36	0.49	Difficult
Organization	Finish	37	0.24	Difficult
Organization	Finish	38	0.64	Very difficult
Organization	Evaluate	39	0.79	Very difficult
Organization	Evaluate	40	0.62	Very difficult
Generation	Designing	41	0.08	Difficult
Generation	Designing	42	0.66	Very difficult
Generation	Compile	43	0.00	Difficult
Generation	Compile	44	-0.13	Easy

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

HOTS level	Indicator	No. Question	Measure Logit	Interpretation
Generation	Make	45	0.39	Difficult
Generation	Make	46	-0.27	Easy
Generation	Produce	47	0.17	Difficult
Generation	Produce	48	0.77	Very difficult
Generation	Develop	49	-0.36	Easy
Generation	Develop	50	0.02	Difficult

The level of ability in the person measure is sorted from highest to lowest and can be seen from the measure. Based on the results of the analysis of individual ability levels at the highest measure to the lowest measure, we found results with a logit of 2.30 to -2.57.

The results show that students who fall into the category of high-level thinking skills are known to number 13 with a percentage of 26%, then students who fall into the category of medium level thinking skills number 143 with a percentage of 286% and students who fall into the category low level thinking skills numbered 148 with a percentage of 296%. Based on the results that have been described, the total percentage of students' high, medium and low-level thinking skills is 608%.

Reliability (Cronbach's alpha, person reliability, and item reliability) and differentiating power (separation)

Reliability is a tool used to measure the constancy (consistency) of measurements and measuring instruments that are reliable (reliable). Reliability can be considered consistent if measurements are repeated and the results obtained remain the same. Test instrument reliability analysis was carried out using the Winsteps program. The Winsteps program can provide information on instrument reliability, namely reliability based on person /test (person separation index), reliability based on items (item separation index), and Cronbach's alpha value, namely the interaction between person and item (Amirrudin, 2021; Taber, 2018). As for the steps that can be taken to display these values, click Output Tables, then click Table 3.1 Summary Statistics. Meanwhile, the output from the Summary Statistics results on the instruments used can be seen in the analysis results (Figure 2).

SUMMARY OF 305 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	20.0	50.0	-.45	.31	1.00	-.1	1.01	.0
S.D.	6.1	.0	.59	.03	.09	1.0	.12	1.0
MAX.	45.0	50.0	2.30	.53	1.31	4.0	1.71	4.1
MIN.	4.0	50.0	-2.57	.29	.81	-2.9	.79	-2.8

REAL RMSE	.32	TRUE SD	.50	SEPARATION 1.56	Person RELIABILITY	.71
MODEL RMSE	.31	TRUE SD	.50	SEPARATION 1.60	Person RELIABILITY	.72
S.E. OF Person MEAN = .03						

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .71

SUMMARY OF 50 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	122.2	305.0	.00	.12	1.00	-.1	1.01	.0
S.D.	35.4	.0	.53	.01	.04	1.1	.06	1.1
MAX.	222.0	305.0	.79	.14	1.09	2.1	1.18	2.5
MIN.	73.0	305.0	-1.50	.12	.90	-3.4	.87	-3.3

REAL RMSE	.13	TRUE SD	.51	SEPARATION 4.08	Item RELIABILITY	.94
MODEL RMSE	.12	TRUE SD	.52	SEPARATION 4.12	Item RELIABILITY	.94
S.E. OF Item MEAN = .08						

Figure 2. Output summary statistics on Winsteps

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

The Summary Statistics output in Winsteps shows the quality of the instruments (items) and respondents (persons) in answering the given instruments. Based on the results obtained, the Summary Statistics Output in Winsteps shows that there is interaction between items and people. Meanwhile the results From Cronbach's alpha, a value of 0.71 was obtained, which means that the interaction between the person and the item or statement item as a whole is good. Interpretation of Cronbach's alpha values can be seen in (Table 4).

Table 4. Interpretation of reliability tests based on *Cronbach alpha values*

Score Criteria	Reliability Level
> 0.8	Very good
0.7 – 0.8	Good
0.6 - 0.7	Enough
0.5 - 0.6	Bad
< 0.5	Bad

The results obtained for the person reliability value were 0.71 and for item reliability it was 0.94. Based on the interpretation of the reliability test above in (Table 4), it can be said that person reliability or consistency of answers from students is included in the good category, meanwhile item reliability is included in the very good category which shows the quality of the items in the instrument which are reliable (Emerson, 2019 ; Park, 2021). Categories of item reliability values based on the Rasch model can be seen in (Table 5).

Table 5. Rasch model item reliability value categories

Score Criteria	Reliability Level
> 0.94	Special
0.91 - 0.94	Very good
0.81 - 0.90	Good
0.67 - 0.80	Enough
< 0.67	Weak

The grouping of people and items or items can be known from the separation value. The greater the separation value, the better the quality of the instrument in terms of overall respondents and items because it can identify groups of respondents and groups of items. The value for person separation is 1.56, so the H value is 2.41. This shows that the person group can be divided into two groups, while for item separation a value of 4.08 is obtained, so the H value is 5.77. This shows that groups of items or questions can be divided into five groups.

Based on the results of the analysis above, it can be said that the instrument for developing high-level thinking skills questions using the Stahl and Murphy Taxonomy in the excretory system material studied has high reliability with the results of the reliability values summarized in (Table 6).

Table 6. Reliability results

Information	Reliability	Category
Cronbach's Alpha	0.71	Good
Person reliability	0.71	Quite good
Item reliability	0.94	Very good

Based on (Table 5) the results show that there is a match between the items on the instrument and the person (respondent). This is supported by obtaining a Cronbach's alpha value of 0.71 which is included in the "good" category, and for the consistency of answers from students (person reliability) which is included in the "fairly good" category, as well as the quality of the items for each item (item reliability) . falls into the "very good" category.

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

CONCLUSION

Based on the results of the analysis that has been described, it can be concluded that the instrument that has been developed regarding the development of high-level thinking skills using Stahl and Murphy's taxonomy on excretory system material is said to be feasible or "valid" and "reliable". This feasibility can be seen from the results of the analysis that has been carried out, namely, the validity analysis shows that 50 questions are fit or acceptable. Students fall into the categories of high, medium and low-level thinking skills respectively, namely around 13, 143 and 148, which means that it can be said that the high-level thinking skills of 8th grade students at Junior High School 2 Sintang are still relatively low.

Reliability analysis shows a person reliability value of 0.71, which means the consistency of students' answers is quite good and item reliability is 0.94, indicating the quality of the items is very good. Cronbach's alpha value is 0.71, which means that the interaction between person and item or item as a whole is good. Based on these results, this instrument can be used to determine the high-level thinking skills of 8th grade students at Junior High School 2 Sintang.

REFERENCES

Afandi, A., & Ningsih, K. (2020). Taxonomy and learning model of higher order thinking. Untan Press.

Amirrudin, M., Nasution, K., & Supahar, S. (2021). Effect of variability on Cronbach alpha reliability in research practice. *Journal of Mathematics, Statistics and Computing*, *17*(2), 223-230.

Azizah, A., & Wahyuningsih, S. (2020). Use of the Rasch model for analysis of test instruments in actuarial mathematics courses. *Journal of Mathematics Education* , *3*(1), 45-50.

Darry, F. W., Fitri, R., & Oktafianti, O. (2021). 4c skills-oriented worksheet in biology for students in senior high school (defining phase). *Atrium Journal of Biology Education*, *6*(3), 220-225.

Emerson, R. W. (2019). Cronbach's alpha explained. *Journal of Visual Impairment & Blindness*, *113* (3), 327-328.

Fanani., & Zainal. (2018). Strategy for developing higher order thinking skills (hots) questions in the 2013 curriculum. *Journal of Islamic Religious Education*, *2*(1), 57-76.

Fitriani, H., Asy'ari, M., Zubaidah, S., & Mahanal, S. (2019). Exploring the prospective teachers' critical thinking and critical analysis skills. *Indonesian Journal of Science Education*, *8*(3), 379-390.

Ichsan, I. Z., Sigit, D. V., Miarsyah, M., Ali, A., Arif, W. P., & Prayitno, T. A. (2019). HOTS-AEP: Higher order thinking skills from elementary to master students in environmental learning. *European Journal of Educational Research*, *8*(4), 935-942.

Irawati, T. N. (2018). Analysis of junior high school students' high-level thinking abilities in solving mathematical problem-solving questions on integer material. *Gammath Journal*, *3*(2), 1–7.

Jaenudin, R., Chotimah, U., Farida, F., & Syarifuddin, S. (2020). Student development zone: higher order thinking skills (hots) in critical thinking orientation. *International Journal of Multicultural and Multireligious Understanding*, *7*(9), 11-19.

Koman, I., Bayrakceken, S., Oktay, O., & Canpolat, N. (2023). Evaluation of science teachers' exam and high school entrance exam science questions based on the revised Bloom taxonomy. *Science Insights Education Frontiers*, *16*(2), 2505-2536.

Mahanal, S., Zubaidah, S., Sumiati, I. D., Sari, T. M., & Ismirawati, N. (2019). RICOSRE: A learning model to develop critical thinking skills for students with different academic abilities. *International Journal of Instruction*, *12*(2), 417-434.

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

Markandan, N., & Osman, K. (2023). Effects of ME-CoT teaching module on students' biology achievement, computational thinking, and metacognitive awareness. *Creative Education*, *14*(2), 288-323.

Nugroho, A. T., Karyanto, P., & Sugiharto, B. (2022). Profile of critical thinking ability of high school students on animalia material during hybrid learning in the pandemic era. *Journal of Science Education Research*, *8*(6), 2635-2640.

Riyani, F., & Siregar, Z. A. (2022). Application of jigsaw cooperative learning model to improve student's learning outcomes in human excretion system materials in private high school Kemala Bhayangkari-2 Rantauprapat. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, *5*(1), 4327-4334.

Rosdiana, S., & Pahlevi, T. (2020). Development of assessment instruments based on higher thinking skills (hots) in archives subjects majoring in OTKP at SMKN 2 Sooko Mojokerto. *Journal of Office Administration Education*, *8*(1), 82-95.

Sari, E. N. F. T., Amin, M., Hudha, A. M., Fatmawati, D., & Fauzi, A. (2021). Development of HOTS-based biology learning documents using ADDIE Model. *Research and Development in Education (RaDEn)*, *1*(2), 61-70.

Saputri, A. C., Rinanto, Y., & Prasetyanti, N. M. (2019). Improving students' critical thinking skills in cell-metabolism learning using stimulating higher order thinking skills model. *International Journal of Instruction*, *12*(1), 327-342.

Sayekti, I. C. (2019). Analysis of the nature of science in class IV student books, sub-theme 1, theme 3, 2013 curriculum. *Basic Education Profession*, *1*(2), 129-144.

Setiawan, J., Sudrajat, A., & Kumalasari, D. (2021). Development of higher order thinking skill assessment instruments in learning indonesian history. *International Journal of Evaluation and Research in Education*, *10*(2), 545-552.

Shafie, H., Majid, F. A., & Ismail, I. S. (2019). Technological pedagogical content knowledge (TPACK) in teaching 21st century skills in the 21st century classroom. *Asian Journal of University Education*, *15*(3), 24-33.

Siburian, J., Corebima, A. D., & Saptasari, M. (2019). The correlation between critical and creative thinking skills on cognitive learning results. *Eurasian Journal of Educational Research*, *19*(81), 99-114.

Susantini, E., Isnawati, & Raharjo. (2022). HOTS-link mobile learning application: Enabling biology pre-service teachers to devise HOTS-based lesson plans. *Journal of Science Education and Technology*, *31*(6), 783-794.

Susilowati, N. E., Muslim, M., Efendi, R., & Samsudin, A. (2022). PISA 2021 creative thinking instrument for students: Physics teachers' perceptions. *Indonesian Journal of Science and Mathematics Education*, *5*(2), 194-209.

Suwono, H., Permana, T., Saefi, M., & Fachrunnisa, R. (2023). The problem-based learning (PBL) of biology for promoting health literacy in secondary school students. *Journal of Biological Education*, *57*(1), 230-244.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273-1296.

Uddin, M.R. (2021). The process of development and validation of a test to measure the critical thinking skills of environmental education among secondary level students of Bangladesh. *International Journal of Educational Research Review*, *6*(3), 264-274.

Park, H. (2021). Reliability using Cronbach alpha in sample surveys. *The Korean Journal of Applied Statistics*, *34*(1), 1-8.

Purnami, W., Ashadi, A., Suranto, S., Sarwanto, S., Sumintono, B., & Wahyu, Y. (2021). Investigation of person ability and item fit instruments of eco critical thinking skills in basic science concept materials for elementary pre-service teachers. *Indonesian Journal of Science Education*, *10*(1), 127-137.

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260

Wening, M.H., & Santosa, A.B. (2020). School principal leadership strategies in facing the digital era 4.0. *Journal of Educational Management, Leadership, and Supervision*, *5*(1), 56–64.

Widiawati, L., Joyoatmojo, S., & Sudiyanto, S. (2018). Higher order thinking skills as effect of problem based learning in the 21st century learning. *International Journal of Multicultural and Multireligious Understanding*, *5*(3), 96–105.

Wihardjo, R., Putri, C. A. A, Putri, A. L. C. A., & Ramadhan, S. (2021). Developing critical thinking assessment instruments on the learning of environmental education. *Review of International Geographical Education Online*, *11*(7), 2485-2492

Wikanta, W., & Susilo, H. (2022). Higher order thinking skills achievement for biology education students in case-based biochemistry learning. *International Journal of Instruction*, *15*(4), 835-854.

Yusuf, I., Widyaningsih, S. W., Prasetyo, Z. K., & Istiyono, E. (2021, March). Blended learning: Its effect towards higher order thinking skills (HOTS). *Journal of Physics: Conference Series*. *1832*(1), 012039.

Zain, F.M., Sailin, S.N., & Mahmor, N.A. (2022). Promoting higher order thinking skills among pre-service teachers through group-based flipped learning. *International Journal of Instructions*, *15*(3), 519-542.

Zhou, Y., Gan, L., Chen, J., Wijaya, T.T., & Li, Y. (2023). Development and validation of a higher-order thinking skills assessment scale for pre-service teachers. *Thinking Skills and Creativity*, *48*(21), 101272.

Acknowledgment

Researcher would like to thank the university which funded this research and the participants who were involved in this research.

Authors' Note

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

How to Cite this Article

Camila, C. G., Afandi, A., Tenriawaru, A. B., Artika, W., & Siregar, N. (2023). Development of higher order thinking skill questions using Stahl and Murphy's taxonomy on excretion system topic. *Assimilation: Indonesian Journal of Biology Education*, *6*(2), 97-108.

DOI: https://doi.org/10.17509/aijbe.v6i2.60632
e-ISSN 2621-7260