JPII 10 (1) (2021) 35-47

Jurnal Pendidikan IPA Indonesia

http://journal.unnes.ac.id/index.php/jpii

SELF-REFLECTION ASSESSMENT IN VERTEBRATE ZOOLOGY (SRAVZ) USING RASCH ANALYSIS

W. L. Yuhanna[1,2], M. H. I. Al Muhdhar[*3], A. Gofur[4], Z. Hassan[5]

[1,3,4]Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang
[2]Department of Biology Education, Faculty of Teacher Training and Education, Universitas PGRI Madiun
[5]School of Education, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia

DOI: 10.15294/jpii.v10i1.25603

Accepted: August 7th 2020. Approved: March 25th 2021. Published: March 31st 2021

ABSTRACT

Instruments that are valid, reliable, and have high consistency are needed to measure students' self-reflection. The Self-Reflection Assessment in Vertebrate Zoology (SRAVZ) was developed to explore students' self-reflection and abilities in the vertebrate zoology course. It is essential to test the instrument's validity before measuring students' abilities so that data bias does not occur. This study aims to determine the validity, whether the items are fit or misfit, and the difficulty level of SRAVZ items. SRAVZ is developed by ADDIE (Analysis, Design, Development, Implementation, Evaluation). The SRAVZ consists of 24 items tested on 135 students who have taken the vertebrate zoology course. Analysis of the Rasch model using Winstep version 4.5.2. The Rasch model shows the item reliability value at 0.97. The Cronbach alpha value at 0.94 with PTMEA Corr shows a positive value, unidimensional 48.1%. The separation index of 5.6 means that the level of grouping the items is very good. The mean square infit for SRAVZ was 0.59-1.96, and the mean square outfit value is 0.59-2.16. Data analysis shows that 24 SRAVZ items have 22 fit items and two misfit items (S3 and S5). Item numbers S3 and S5 must be excluded from the SRAVZ construction. Total items used to measure students' self-reflection in the vertebrate zoology course were 22 items. The most difficult item is S3, and the easiest item is S6. Thus, the data indicate that the valid and reliable SRAVZ is in the good, effective, and high level of consistency category.

Keywords: self reflection; vertebrate zoology; rasch

INTRODUCTION

Vertebrate Zoology is one of the compulsory subjects that must be taken by undergraduate biology education students. This course contains basic taxonomy material, characteristics, and classifications of vertebrates, pisces, amphibians, reptiles, aves, and mammals. The vertebrate zoology class is conducted by learning in class, doing practicum, and joining field study. The teaching and learning process uses various interesting and relevant methods. The practicum is part of the credit system, which is held around 4-6 times depends on the material. Moreover, it serves to support the theories given by lecturers, to develop students' scientific attitudes, and to improve students' long-term memory about the concept of vertebrate zoology.

The success of students in taking vertebrate zoology courses is influenced by various aspects: lecture planning, lecture implementation, material understanding, practicum, and lecture evaluation. The five aspects must be developed so that students will understand the concept of vertebrate zoology, then hopefully, they can apply it in real life. However, students often do not pay attention to those five aspects, and it causes failure in taking the course. The ability of students to independently measure success in taking a course is highly needed to form self-regulation and self-efficacy. Students who can assess themselves well will learn optimally and increase their competence effective (Chen & Lin, 2018; Yan, 2020).

***Correspondence Address**
E-mail: mimien.henie.fmipa@um.ac.id

Self-reflection and the development of insight are needed in the learning process to change behavior and mindset (Chen & Lin, 2018; Chen et al., 2019; Yan et al., 2020b). Self-reflection covers various aspects of planning and creating the learning process (Yuhanna & Retno, 2016; Diefes-Dux & Castro, 2018), influencing the final result. Thus, students can understand the results after attending lectures and analyze the difficult concepts. Self-reflection is related to how and what students have learned leads to academic performance improvement (Lew & Schmidt, 2011; Cavilla, 2017) . It also provides information for lecturers to improve quality and develop the course.

Self-reflection is one part of psychometric measurements. Psychometric measurements are susceptible to the participant's conditions. Currently, self-reflection measurements are carried out in general, without knowing the internal and external conditions of students—likewise, the measurement of self-reflection in vertebrate zoology learning. Wrong statement items can trigger data bias (Boone et al., 2011). Self-reflection measurement data in a course must accommodate students' abilities in understanding the material discussed and the lecture model used.

Self-reflection in the vertebrate zoology course needs to be done to determine the strengths and weaknesses of students. Self-reflection instruments for vertebrate zoology courses have not been developed. Measuring self-reflection using instruments with general statements cannot accommodate students' learning abilities and conditions. A good instrument is currently able to measure what should be measured (Widhiarso & Sumintono, 2016). Self-reflection instruments following the characteristics of the vertebrae zoology course are needed in independent learning during the COVID-19 pandemic. A good instrument is an instrument that can explore the intangible aspects to bring up the ability to be measured.

The self-reflection measurement instrument is needed to measure the students' ability to learn vertebrate zoology. However, self-reflection questionnaires that specialized in vertebrate zoology have not been developed much yet. In measuring learning success, lecturers measure not only their learning outcomes but also their ability. Measurement of self-reflection in the vertebrate zoology course is needed. Ability measurement starting from planning, learning implementation, concept understanding, practicum implementation, and learning evaluation is expected to make students understand the contents of vertebrate zoology in the long-term memory stage. Thus, the researcher developed this instrument based on the five (5) constructs of lecturing. The five constructs were arranged into a complete instrument.

The evaluation of students' success in taking a course needs to be measured with appropriate and valid instruments. A valid instrument is the first step to produce accurate data (Misbah et al., 2018; Pichardo et al., 2018). A reliable instrument is needed to measure a variable. A reliable instrument must have good validity and reliability to be used latently and comprehensively (Ariffin et al., 2010; Mohamad et al., 2015; Arnold et al., 2018). The reliability of the instrument is expected to be able to explore students' abilities based on the attributes measured (Susongko, 2016). Testing the validity and reliability of the instrument is carried out as a form of a pilot test (Yasin et al., 2015) to test the reliability of the instrument in measuring student's overall ability.

Gap analysis in this study focuses on self-reflection instruments to measure students' success holistically in taking Vertebrate Zoology. Specific and valid instruments for measuring self-reflection have not been developed, and their reliability and validity have not been tested. Thus, it is necessary to develop and test the validity to produce a reliable self-reflection instrument for Vertebrate Zoology to be used for further research.

The researchers developed the Self-Reflection Assessment of Vertebrate Zoology (SRAVS) consisting of 24 statements from 5 constructs of vertebrate zoology learning activities. The pilot test aims to ensure that the instrument is relevant to what will be measured and the person to be measured. The instrument reliability can be seen from the items' completeness based on the students' abilities and the clarity of the item test's meaning. Validation is the collection of evidence to provide a scientific basis of test scores interpretation (Yang et al., 2018). Also, validation is the process of generating and interpreting evidence to conclude that the tool used matches the attributes that will be measured (Peeters & Martin, 2017). Validity is related to the appropriateness of the content, representation, and technical quality (Susongko, 2016). The validity of a measuring instrument is the most fundamental consideration in developing and evaluating tests (Harpe, 2015). The high validity instruments can reduce the bias of measuring instruments, which results in missing data.

The problems in analyzing the validity of the psychometric instrument using parametric statistical tests were the incompatibility of items with the variables to be measured (Linacre, 2012; Maat & Rosli, 2016), the absence of an analysis of students' ability to answer questions, and the existence of biased items. The validity test of psychometric instruments using statistical analysis was also unable to predict the missing data scores (Garzón Umerenkova et al., 2017). The respondent's answer pattern also cannot be interpreted.

Rasch model can provide a strong guiding framework for those developing new instruments (Boone, 2016). Rasch model can be used in pilot tests to test instrument validity and reliability. Rasch model is a data analysis technique based on the item and person quality (Kudiya et al., 2018; Maseko et al., 2019). Rasch model refers to the level of difficulty of items in determining the quality of people. This technique is different from the usual techniques because it is not based on the number of correct answers. The Rasch model can analyze the suitability of items in instrument development. This method can predict missing data. Respondents' response patterns can also be used as a basis for identifying partial data.

Rasch model is also able to record psychometric components that allow evaluation of several characteristics, such as model fit level, item difficulty level and hierarchy, reliability of person and item, and item function differentials (DIF) (Cupani et al., 2017). The complexity of psychological attributes can also be used as consideration material related to the dimension to be measured (Fisher, 2017).

Besides, objective measurements in social research, for example, in an educational area, must meet five criteria: 1) Producing linear measurements with the same intervals; 2) Summing up estimation process; 3) Identifying incorrect items (misfits) or unusual items (outlier); 4) Overcoming missing data; 5) Producing measurements that do not depend on the parameters studied (Linacre, 2012; Yan et al., 2020a). The advantages of analysis using the Rasch Model are being able to meet five criteria and have the same quality as measurements made in physical dimensions in the field of physics (Widhiarso & Sumintono, 2016).

Moreover, the Rasch model is widely used in social, educational, economic, and scientific research. The instruments for measuring a construct are also mostly done using the Rasch model. The measurement of item validity using the Rasch model has been performed on the Scientific Inquiry Competence instrument (Arnold et al., 2018), Self-Regulation Questionnaire (Garzón Umerenkova et al., 2017; Pichardo et al., 2018), Content Knowledge in biology inventory (Großschedl et al., 2018). Rasch analysis is also used to measure the validity of items from test questions in biology courses (Cupani et al., 2017), evolution (Fiedler et al., 2019), and energy (Herrmann-abell et al., 2018).

The urgency of this research is to produce a valid self-reflection measuring tool in the vertebrate zoology course that can be used to measure the success rate of students in taking the whole course. SRAVZ needs to be tested for validity before producing reliable measuring instruments, appropriate question items, and unbiased items. The validity of the SRAVZ needs to be tested to ensure that this instrument can more accurately and reliably measure the intangible aspects of self-reflection in the vertebrate zoology course. This study also reveals the relationship between items in developing instrument constructs holistically that cannot be analyzed using statistics.

Researchers did not use demographic data person. The conditions of the students learning environment in supporting vertebrate zoology course are also not seen. This research is a preliminary study to test the validity of SRAVZ. The goals of this study are to 1) Identify the validity of the SRAVZ using the Rasch model, 2) Identify the fit and misfit items of SRAVZ, 3) Analyze the level of difficulty of SRAVZ items.

METHODS

This research aims to assess the validity and suitability of SRAVZ items developed by the researchers. SRAVZ development used the ADDIE model (Analyze, Design, Development, Implementation, Evaluation). The research was conducted from March to April 2020. The respondents were 135 students who had taken vertebrate zoology courses from the three Universities.

SRAVZ consists of 5 constructs: lecture planning, learning implementation, material understanding, practicum implementation, and evaluation (Table 1). This questionnaire consists of 24 statements. Two assessment experts previously validated SRAVZ. The questionnaire used a Likert scale (Harpe, 2015) consisting of 5 answer choices, namely "excellent," "very good, "good," "fair," and "poor." SRAVZ composition is as in Table 1.

After data was collected, it was processed using the Rasch model by applying Winstep 4.5.2 software. Rasch model is used as a better way to convert raw scores into ratio scores so that people's abilities can be measured on a ratio scale. Several studies have used Rasch modeling to validate the instruments used (He et al., 2016; Yang et al., 2018).

Table 1. Item Self Reflection Assessment in Zoology Vertebrate (SRAVZ)

Construct	Statements	Code	Answer				
			Excellent (5)	Very Good (4)	Good (3)	Fair (2)	Poor (1)
Lecture Planning	I understood the prerequisite course material (invertebrate zoology) before taking the vertebrate zoology course	S1					
	I understood the Semester Lecture Plan and the course contract of Vertebrate Zoology	S2					
	I had a Vertebrate Zoology reference books based on the lecture plan or lecturer recommendation	S3					
	I planned the learning mechanism of vertebrate zoology before attending the lecture	S4					
	I had a plan and target for the final grade of the vertebrate zoology course	S5					
Learning Implementation	I attended all lectures of Vertebrate Zoology	S6					
	I gave attention to the lecturer's explanation	S7					
	I participated in the presentation	S8					
	I joined the discussion	S9					
Material understanding	I understood the basic materials of taxonomy	S10					
	I understood the characteristics and classifications of vertebrate	S11					
	I understood the material about Pisces	S12					
	I understood the material about Amphibian	S13					
	I understood the material about Reptile	S14					
	I understood the material about Aves	S15					
	I understood the material about Mammals	S16					
	I understood the scientific name of vertebrate animal	S17					
Practicum implementation	I read the practicum direction	S18					
	I understood the tools and materials for practicum	S19					
	I did a practicum of vertebrate zoology	S20					
	I completed the assignments given by lecturers	S21					
Evaluation	I completed the practicum report	S22					
	I prepared and took mid-test	S23					
	I prepared and took the final test	S24					

The validity of the SRAVZ instrument can be seen from the reliability and separation index, unidimensionality, item polarity, fit and misfit items, and items' difficulty level. The range of reliability and Cronbach alpha (α) values as in Table 2 below.

Table 2. Reliability and Cronbach Alfa (α) Score

Reliability and Cronbach Alfa (α) Score	Means
0,9 - 1,0	Very good, effective at a high level of consistency
0,7 - 0,8	Good and Acceptable
0,6 - 0,7	Acceptable
<0,6	The item needs to be refined
0,5	The item needs to be dropped

To check fit and misfit items, Maat and Rosli (2016) explained that the mean square of infit and outfit must be between 0.60-1.40. A logit (measure) value that does not meet these rules is declared a misfit.

RESULTS AND DISCUSSION

SRAVZ validity testing was carried out to produce reliable, specific instruments capable of measuring student ability in taking the vertebrate zoology course. Testing the instrument's validity using the Rasch model can show the level of sensitivity of the instrument in measuring the variable of personal abilities. The results of the preliminary research on the validity of SRAVZ are shown in Table 3.

Table 3. Reliability and Separation of SRAVZ

Criteria	Measurement	
	Item	Person
Mean	0,0	2,20
S.D	0.82	1,38
Reliability	0.97	0,90
Separation index	5,36	2,94
Cronbach's alpha (α)	0,94	
PT MEA Corr	0,47 - 0,76 (positive)	
Raw variance explained by measures	48,1%	

Rasch model can facilitate the development of instruments that provide useful data (Linacre, 2012; Boone & Noltemeyer, 2017). It can provide data that can be used confidently for both descriptive (Arnold et al., 2018; Planinic et al., 2019). The aspects discussed in this validating SRAVZ are reliability (item and person), Cronbach's alpha (α), separation index, item polarity, and unidimensionality.

Table 3 showed that the mean value of the items was 0.00, and the mean person was 2.20. Standard Deviation (SD) items were 0.82, and SD person was 2.20. The reliability item was 0,97, and the person was 0,90. Cronbach's alpha (α) was 0,94. The criteria for the accepted value of α was between 0.71-0.99 (Table 2). This study showed that the α value was 0.94 (Table 3), which indicated that the reliability of the item and person relationship was very good, effective, and had a high consistency.

The separation item index showed a group of difficulty levels. Separation index items showed a value of 5.36 (Table 3), which can be interpreted that the level of difficulty items can be grouped into five. Separation index value > 2.0 indicated that grouping items belonged to good criteria. The separation person index was 2.94, which showed the instrument could be distinguished into three groups based on participants' self-reflection. The separation person index showed the calculation of the instrument's ability to group individuals for various levels based on what was measured.

Polarity items indicate the presence or absence of relationships between items built with a person. The polarity item on the Rasch model can be seen from the Point Measure coefficient (PTMEA Corr). Table 3 showed that from 24 items, the PTMEA Corr values were positive (0,47-0,76). A positive PTMEA Corr value indicates that items can be used to measure in a relevant manner. In contrast, the negative PTMEA Corr value indicates that the item cannot measure the construct, and the item must be abandoned.

Unidimensionality is a measure to evaluate the ability of an instrument to measure the ability to be measured. Unidimensionality is the items of the scale assess one single underlying latent construct (Upegui-Arango et al., 2020). The Rasch model uses Principal Component Analysis from residuals to measure instrument diversity

according to the construct (Boone & Noltemeyer, 2017). Unidimensionality in this study was 48.1% (Table 3). It shows that as many as 48,1% of unidimensionality requirements can be fulfilled.

Data analysis using the Rasch model can determine the instrument's validity (Boone & Scantlebury, 2006). The validation results for the SRAVZ instrument showed that the α value of 0.94 means that the SRAVZ instrument was very good, effective, and had a high level of consistency in terms of the relationship between items and people. The α value of 0.94 fulfilled the validity criteria of instruments with high reliability. The value of the reliability item was 0.97, which indicated that the instrument was sensitive enough to distinguish people with different abilities (Park & Liu, 2019).

The items on SRAVZ showed latent characteristics and can function as valid measurements. Item reliability was 0.97 indicating the instrument was relevant and can be used repeatedly, and it did not depend on the environment observed.

The reliability value of the person in this study was 0.90 (Table 3), which indicated that the consistency of individuals who responded to statements was very good. Person reliability of 0.90 also indicated that there was no problem with the person. Also, respondent answered with a high level of consistency. Instrument reliability analysis using the Rasch model can be used to reduce the possibility of duplicate items in the construct area or aspect to be measured (Herrmann-abell et al., 2018).

The separation index of items showed a value of 5.36, which can be interpreted that the SRAVZ instruments can be grouped into five levels. The separation index of a person was 2,94. This grouping was categorized in detail, and it could measure the components. Separation index values can range from 0 to infinity, and higher values indicate better separation. Item separation indices of 3 or greater are desirable (Van Zile-Tamsen, 2017). In terms of person separation, an index of 1.50 is acceptable, 2.00 is good, and 3.00 is excellent (Widhiarso & Sumintono, 2016). In this study, the separation of items and people can meet the standard criteria in measuring instrument validity.

Besides, PTMEA Corr calculation is critical to be used as a study material in seeing the validity of an instrument (Planinic et al., 2019). The results of the PTMEA Corr calculation showed that 24 item items had a positive value (0,47-0,76). Item polarity was analyzed to identify whether those items function according to positive point correlations (Mayes et al., 2019). Based on the result, the item polarity was all positive and relatively strong to be used as a measurement. Unidimensionality was 48,1%, it is shown that items in SRAVZ can measure self-reflection in the vertebrate zoology course.

Identifying fit and misfit is necessary to determine suitable items for measuring student self-reflection in the vertebrate zoology course. Fit items show the suitability of the statement in measuring student ability. Misfit items represent inappropriate statements if used as measuring instruments and should be removed from construction items in the instrument (Boone, 2020).

Based on the Winstep analysis 4.5.2 version, from 24 questions related to SRAVZ displayed in Table 4. Data showed that there were both fit and misfit instrument items. To check fit and misfit items, Maat and Rosli (2016) explained that the mean square of infit and outfit must be between 0.60-1.4. A logit (measure) value that does not meet these rules is declared as misfit.

Table 4 illustrated the mean square values of the infit and outfit of the 24 items tested. The mean square infit value for SRAVZ was 0.59-1.96, and the mean square outfit value was 0.59-2.16. Identification from Table 4 showed that from 24 SRAVZ items, there were 22 fit items and two misfit items. Misfit items were found on S3 and S5. Misfit items are items with an MNSQ infit value and MNSQ outfit outside the range of 0.60-1.40 (Bond et al., 2020). Infit value S3 was 1,96, and S5 was 1,60. Outfit value S3 was 2,16 and S5 was 1,65. Those two items did not meet the instrument validation requirements, so they should not assess students' self-reflection in the vertebrate zoology course. Item analysis is the best method for controlling the quality of instruments used in measuring psychometric aspects (Sabudin et al., 2018; Chao et al., 2019; Aghekyan, 2020).

Table 4. Fit and Misfit Items of SRAVZ

ITEM	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEA			EXACT MATCH	
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.		EXP	OBS%	EXP%
S3	471	135	1.51	0.12	1.96	6.01	2.16	7.05	A	0.48	0.69	36.10	50.50
S5	575	135	-0.20	0.14	1.60	3.89	1.65	3.76	B	0.47	0.59	54.10	60.60
S1	494	135	1.19	0.12	1.19	1.42	1.39	2.80	C	0.57	0.67	47.40	51.80
S6	620	135	-1.31	0.17	1.31	2.05	1.04	0.25	D	0.47	0.50	62.40	69.70
S20	542	135	0.42	0.13	1.28	2.00	1.19	1.39	E	0.66	0.63	50.40	56.20
S4	490	135	1.24	0.12	1.25	1.83	1.27	2.02	F	0.52	0.67	45.90	51.60
S21	616	135	-1.19	0.17	1.25	1.73	1.04	0.27	G	0.53	0.51	70.70	68.20
S22	577	135	-0.24	0.14	1.18	1.35	1.08	0.59	H	0.60	0.59	63.20	60.70
S2	536	135	0.53	0.13	1.13	0.99	1.14	1.04	I	0.55	0.63	59.40	55.70
S23	612	135	-1.08	0.17	1.07	0.54	1.14	0.72	J	0.51	0.52	68.40	67.10
S24	618	135	-1.25	0.17	1.09	0.65	0.99	0.05	K	0.48	0.50	65.40	68.90
S18	550	135	0.28	0.13	1.00	0.07	0.95	-0.32	L	0.66	0.62	59.40	57.10
S19	560	135	0.09	0.14	1.00	0.06	0.94	-0.43	l	0.65	0.61	61.70	58.60
S8	612	135	-1.08	0.17	0.90	-0.74	0.99	0.04	k	0.52	0.52	71.40	67.10
S17	490	135	1.24	0.12	0.92	-0.58	0.98	-0.15	j	0.65	0.67	51.90	51.60
S9	601	135	-0.79	0.16	0.88	-0.91	0.97	-0.12	i	0.54	0.54	69.90	64.70
S16	575	135	-0.20	0.14	0.76	-1.91	0.67	-2.53	h	0.69	0.59	68.40	60.60
S7	602	135	-0.81	0.16	0.72	-2.29	0.75	-1.46	g	0.60	0.54	73.70	64.80
S11	560	135	0.09	0.14	0.68	-2.69	0.71	-2.33	f	0.70	0.61	69.20	58.60
S15	565	135	0.00	0.14	0.66	-2.85	0.62	-3.35	e	0.73	0.60	72.90	59.00
S13	553	135	0.23	0.14	0.63	-3.24	0.61	-3.48	d	0.76	0.62	72.90	57.90
S10	521	135	0.77	0.13	0.62	-3.38	0.61	-3.48	c	0.71	0.65	67.70	53.90
S12	549	135	0.30	0.13	0.61	-3.69	0.60	-3.47	b	0.76	0.62	70.70	57.00
S14	552	135	0.24	0.14	0.61	-3.54	0.61	-3.49	a	0.76	0.62	72.90	57.70
MEAN	560.00	135.00	0.00	0.14	1.01	0.10	1.00	0.20				62.80	59.60
P.SD	43.10	0.00	0.83	0.02	0.33	2.50	0.36	2.50				10.00	5.70

Note: S.E =Standart Error, MNSQ = Mean Squares, PT MEA Corr. = Point Measure Correlation

Based on Table 4, data fit items were 22 statements, as follows: (S1, S6, S20, S4, S21, S22, S2, S23, S24, S18, S19, S8, S17, S9, S16, S7, S11, S15, S13, S10, S12, S14). The range of MNSQ infit values for item fit was 0.60 - 1.25. The small MNSQ infit score shows the value of "suitability" items in measuring students' self-reflection after taking vertebrate zoology course. Item fit shows that the item statement is able to explore individual student abilities according to their respective conditions (Pleasence & Balmer, 2019) This item can sort students into groups of students with specific abilities (Lambri et al., 2019).

The interesting fit data items in Table 4 were six statements in the last row (S11, S15, S13, S10, S12, S14). The items of this data were the best in measuring and grouping students according to their abilities. Also, those six statements were part of the understanding aspects of vertebrate zoological material which consist of statements related to the material characteristics and classification of vertebrates (S11), Aves (S15), amphibians (S13), basic taxonomic basis (S10), Pisces (S12), and reptiles (S14). The ability of individuals in self-reflection related to understanding concepts can be used as study material for the lecture to improve learning methods and mastery of the material. This item was expected to measure the level of understanding of vertebrate zoological material based on self-reflection of students, which can be compared with the result of students' learning outcomes on each theme. Based on fit and misfit items, there were 22 fit items and two misfit items (Table 5). Table 5 showed about item fit and misfit based on the construct.

Table 5. Item Fit and Misfit based on the Construct

Construct	Fit items	Misfit items
Lecture Planning	I understood the prerequisite course material (invertebrate zoology) before taking the vertebrate zoology course (S1)	I had Vertebrate Zoology reference books based on the lecture plan or lecturer recommendation (S3)
	I understood the Semester Lecture Plan and the course contract of Vertebrate Zoology (S2)	I had a plan and target for the final grade of the vertebrate zoology course (S5)
Learning Implementation	I planned the learning mechanism of vertebrate zoology before attending the lecture (S4)	-
	I attended all the lectures of Vertebrate Zoology (S6)	-
	I gave attention to the lecturer's explanation (S7)	-
	I participated in the presentation (S8)	-
	I joined the discussion (S9)	-
Material understanding	I understood the basic materials of taxonomy (S10)	-
	I understood the characteristics and classifications of vertebrate (S11)	-
	I understood the material about Pisces (S12)	-
	I understood the material about Amphibian (S13)	-
	I understood the material about Reptile (S14)	-
	I understood the material about Aves (S15)	-
	I understood the material about Mammals (S16)	-
	I understood the scientific name of vertebrate animals (S17)	-
Practicum implementation	I read the practicum direction (S18)	-
	I understood the tools and materials for practicum (S19)	-
	I did a practicum of vertebrate zoology (S20)	-
	I completed the assignments given by lecturers (S21)	-
Evaluation	I completed the practicum report (S22)	-
	I prepared and took mid-test (S23)	-
	I prepared and took the final test (S24)	-
Total	**22 items**	**2 items**

Based on the study results (Table 5), there were two misfit items on the SRAVZ instruments: S3 and S5. Misfit item (S3) was part of the ***Lecture Planning*** construct which contains a statement that *"I had a vertebrate zoology reference book based on the lecture plan or lecturer recommendation"*. In this statement from 135 respondents showed that 19% of respondents answered "Excellent", 41% answered "very good", 22% answered "good", 8% answered "fair", and 10% answered "poor". This statement was in the misfit zone with the highest difficulty level. It indicated that students did not use reference books based on the lecturer's direction and the lecture plan. Reference books were

a vital component that must be possessed by students to support their learning. This data showed that the planning aspects of lectures on the availability of books (S3) items did not need to be asked in measuring the reflection of vertebrate zoology course.

Furthermore, the misfit item S5 contained a statement that *"I planned and had a final grade target for vertebrate zoology courses"*. The number of respondents who answered the statement "excellent" was 48%, "very good" was 35%, "good" was 13%, "fair" was 2% and "poor" 1%. This data showed that most students had a plan and an overview of the target score in the vertebrate zoology courses. The target value showed that students think futuristic by preparing courses. This aspect was necessary because the target encouraged motivation and enthusiasm for learning. Students can also manage their learning systems according to their desired targets. Thus, this S5 statement should not be asked again in measuring the reflection of students in taking vertebrate zoology courses.

The level of difficulty of the items can be seen in Figure 1 and Table 5. Table 5 showed that the level of difficulty of the questions from the most difficult to the easiest. The Wright map (figure 1) presents that the SRAVZ items are plotted on a vertical line representing the logit scale of the instrument. Items plotted at the base of the Wright map are items that are easier to agree with than items plotted toward the top of the Wright Map (Brann et al., 2020). The Wright map showed that the most difficult item was S3 (logit: 1.51), which came from the lecture planning construct. Whereas the easiest item was S6 (logit: -1,31) which came from the lecture implementation construct. The level of difficulty of the questions can determine the variation and level of instruments in measuring the ability of people.

Figure 1 showed that the difficulty level of SRAVZ items. The easier items were placed at the lower end of the map, while difficult items were located at the higher end. The most difficult item to approve was item S3. Item S3 was part of the ***Lecture Planning*** construct related to the statements about the availability of reference books. This item was difficult for the person to approve because it is influenced by material availability, reading habits, and learning styles. In this item, the instructor needs to explore more in-depth the readiness of students before attending the vertebrate zoology course. The respondents' selection in this pilot test consists of people who have different learning styles and a different gender. The existence of misfit items and the difficulty of these items provide a reference for researchers that the book is not the main component of learning material for students with particular learning styles. Besides having a high level of difficulty, S3 is a misfit one and cannot measure student self-reflection in taking vertebrate zoology courses.

The most easily agreed item by the person was item S6. This item contained the statement *"I was fully followed vertebrate zoology lectures"*. This item was effortless to understand because it was related to the completeness and discipline in attending vertebrate zoology course. The aspect of full attendance in attending lectures becomes a vital component, so students automatically understand this item quickly. Respondents from various learning styles and gender had the same perception in interpreting item S6.

In this aspect, the most difficult item to approve was S3 with the statement "I had a Vertebrate Zoology reference book based on the Lecture plan or lecturer recommendation." The easiest item to approve was S5 with the statement "I had a plan of a final grade target for vertebrate zoology courses". Items in this construct provide a picture of self-reflection related to preparation before taking vertebrate zoology courses which are sometimes not realized by students. Item S5 was easiest to approve because basically, students have a vision of the final score obtained before starting lectures in the vertebrate zoology course.

Figure 1 shows that the level of difficulty items in each construct is presented. The ***Lecture Planning*** construct of the lecture consisted of 5 statements (S1, S2, S3, S4, S5). In this aspect, the most difficult item to approve was S3 with the statement "I had a Vertebrate Zoology reference book based on the Lecture plan or lecturer recommendation." The easiest item to approve was S5 with the statement "I had a plan of a final grade target for vertebrate zoology courses." Items in this construct provide a picture of self-reflection related to preparation before taking vertebrate zoology courses which are sometimes not realized by students. Item S5 was easiest to approve because basically, students have a vision of the final score obtained before starting lectures in the vertebrate zoology course.

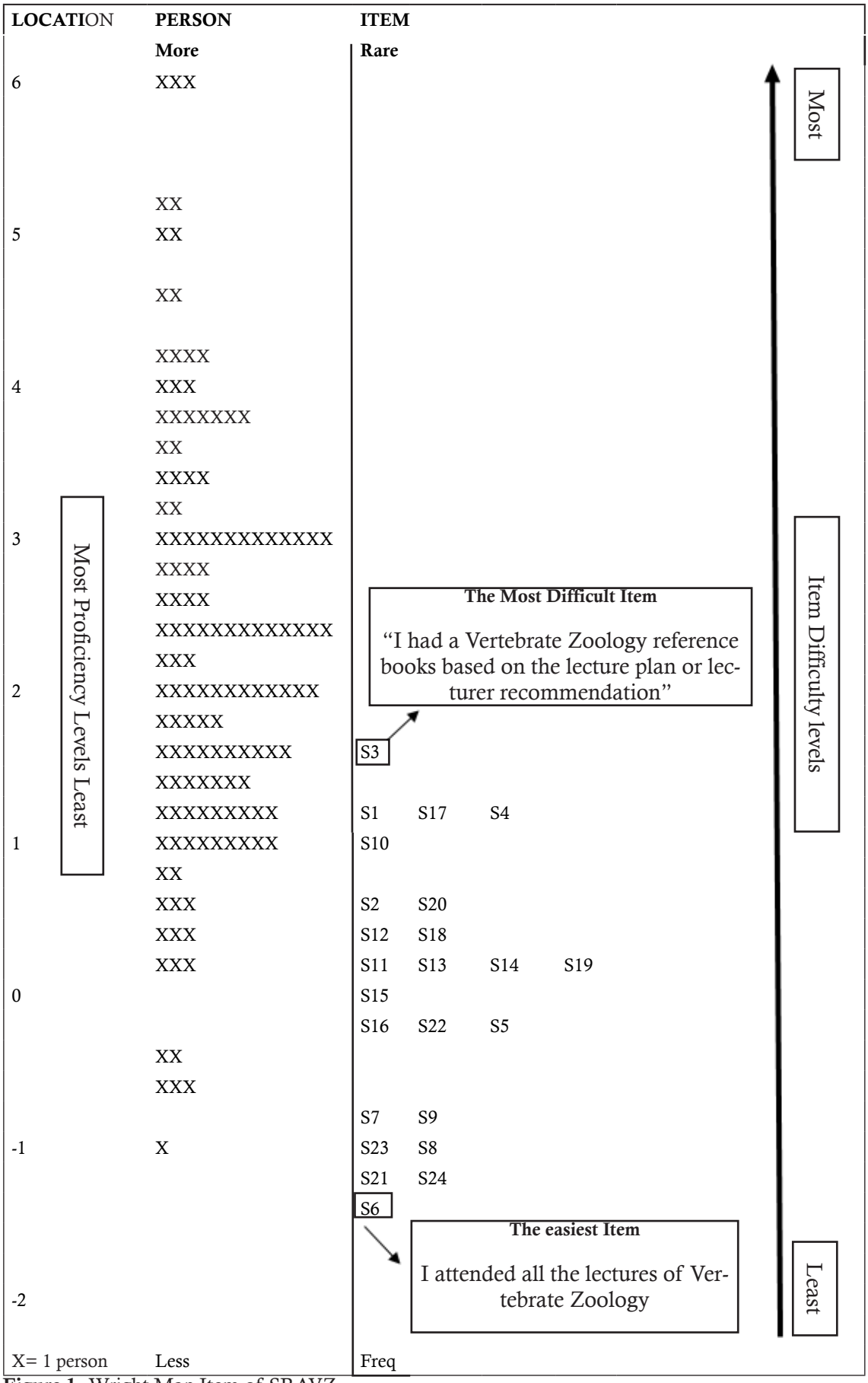

Figure 1. Wright Map Item of SRAVZ

The construct of the ***implementation of vertebrate zoology lectures*** consisted of 4 statements (S6, S7, S8, S9). The most difficult item to approve was S9 (logit = -0.79) with the statement "I joined the discussion." This item was less approved because some of the lecture processes did not implement the discussion method. Thus, students felt doubt in answering this item. Besides, respondents with different learning styles and gender had their views in interpreting this item number. While the easiest item to approve was S6 (logit value = -1.31) with the statement "I was fully followed Vertebrate Zoology lectures." This item was easiest in constructing lecture implementation, and it was also the easiest item out of 24 SRAVZ items.

The construct of u***nderstanding vertebrate zoological material*** consisted of 8 statements (S10, S11, S12, S13, S14, S15, S16, S17). Those statements were aimed to determine students' self-reflection in understanding vertebrate zoological material. The most difficult item was S17 (logit = 1.24) with the statement "I understood the scientific name of vertebrate animals" while the easiest item was the S16 statement (logit = -0.20) with the statement "I understood the mammal material." The level of understanding material from the most difficult to the easiest is as follows: the understanding of the scientific name of vertebrate animals, the basics of taxonomy, Pisces, reptiles, amphibians, characteristics, and classification of vertebrates, Aves, and mammals. This item can be used as study material for lecturers and students in designing effective and efficient learning.

There are three statements in the ***practicum implementation*** construct (S18, S19, S20). Practically, lectures on the vertebrate zoology course done in two activities: doing direct learning (presentation and discussion of the material in classroom) and conducting practicum. The most difficult practicum implementation item was S20 (logit = 0.42) with the statement "I did vertebrate zoology practicum." This item showed that the respondents did not understand the essence of practicum. This item was less specific related to the purpose of conducting practicum. The easiest item was Q19 (logit = 0.09) with the statement "I understood the tools and practical material." Students, as respondents had a good understanding of interpreting this item. Mastery of tools and materials was a competency that must be possessed by students in practicing vertebrate zoology practicum.

The ***evaluation*** construct consisted of 4 statements (S21, S22, S23, S24). The most difficult item in this aspect was S22 (logit -0.24) with the statement "I completed the practicum report." Students were difficult to accept and understand this item. While the easiest item was S24 (logit = -1.25) with the statement "I prepared and took the final semester exams." The final semester exam is a compulsory component for students in taking the vertebrate zoology course, so students easily understood this item.

The finding of this study was the existence of an instrument to measure the success of students in taking the vertebrate zoology course. The SRAVZ contains holistic and specific questions covering lecture planning, learning implementation, material understanding, practicum implementation, and evaluation. Students can use it to measure their ability to learn regardless of gender and learning style. SRAVZ has been tested and analyzed, and it showed as a valid and reliable instrument. Further improvements and validations are needed so that the instrument can become widely applicable (Yang et al., 2018). In general, SRAVZ can be used as an instrument in measuring student's self-reflection without bias.

CONCLUSION

SRAVZ is an instrument developed specifically to explore students' self-reflection and abilities in the vertebrate zoology courses that meet the validity criteria using the Rasch model. The validity test of the instrument is crucial before measuring the student's ability. Based on the results and discussion, it can be concluded that 1) The item reliability value was 0.97, Cronbach alpha value was 0.94, separation items were 5.36, separating person was 2,94, PTMEA Corr values were positive, Unidimensionality was 48,1%. 2) Rasch's analysis showed that from 24 statements on the SRAVZ, there were 22 fits and two misfits (S3, S5) items. The last two items must be discarded in measuring students' self-reflection in studying vertebrate zoology. 3) The level of difficulty of the problem showed that the most difficult statement was S3 (I had a vertebrate zoology reference book based on the recommendation of the lecturer and the lecture plan). The easiest item was S6 (I attended all vertebrate zoology lectures). The SRAVZ instrument was expected to measure students' self-reflection in establishing vertebrate zoology courses.

ACKNOWLEDGEMENTS

The author would like to thank State Universitas Negeri Malang for providing funding for this research with number 4.3.414/UN32.14.1/LT/2020. Secondly, Mrs. Nurmala Hindun, Mr. Riyanto, Mrs. Lely Mardiyana, Mrs. Agita Risma Nurhikmawati and Mrs. Sri Lestari for technical assistance in this study.

REFERENCES

Aghekyan, R. (2020). Validation of the SIEVEA instrument using the Rasch analysis. *International Journal of Educational Research*, *103*, 101619.

Ariffin, S. R., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability multiple intelligent item using rasch measurement model. *Procedia-Social and Behavioral Sciences*, *9*, 729-733.

Arnold, J. C., Boone, W. J., Kremer, K., & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch measurement techniques. *Education Sciences*, *8*(4), 184.

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, *15*(4), rm4.

Boone, W. J. (2020). Rasch Basics for the Novice. In *Rasch Measurement* (pp. 9-30). Springer, Singapore.

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1416898.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253-269.

Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education*, *95*(2), 258-280.

Brann, K. L., Boone, W. J., Splett, J. W., Clemons, C., & Bidwell, S. L. (2020). Development of the School Mental Health Self-Efficacy Teacher Survey Using Rasch Analysis. *Journal of Psychoeducational Assessment*, 0734282920947504.

Cavilla, D. (2017). The effects of student reflection on academic performance and motivation. *SAGE Open*, *7*(3), 2158244017733790.

Chao, J., Siu, A. M., Leung, O., Lo, A., Chu, M., Lee, W. K., ... & Chien, C. W. (2019). Chinese version of the Recovery Self-Assessment scale: psychometric evidence from Rasch analysis and reliability estimates. *Journal of Mental Health*, *28*(2), 206-212.

Chen, F. F., Chen, S. Y., & Pai, H. C. (2019). Self-reflection and critical thinking: the influence of professional qualifications on registered nurses. *Contemporary nurse*, *55*(1), 59-70.

Chen, Y. H., & Lin, Y. J. (2018). Validation of the short self-regulation questionnaire for Taiwanese college students (TSSRQ). *Frontiers in Psychology*, *9*, 259.

Cupani, M., Zamparella, T. C., Piumatti, G., & Vinculado, G. (2017). Development of an Item Bank for the Assessment of Knowledge on Biology in Argentine University Students. *Journal of applied measurement*, *18*(3), 360-369.

Diefes-Dux, H. A., & Castro, L. M. C. (2018, October). Student reflection to improve access to standards-based grading feedback. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.

Fiedler, D., Sbeglia, G. C., Nehm, R. H., & Harms, U. (2019). How strongly does statistical reasoning influence knowledge and acceptance of evolution?. *Journal of Research in Science Teaching*, *56*(9), 1183-1206.

Fisher Jr, W. P. (2017). Suggestions for rethinking validation. *Measurement: Interdisciplinary Research and Perspectives*, *15*(2), 86-90.

Garzón Umerenkova, A., de la Fuente Arias, J., Martínez-Vicente, J. M., Zapata Sevillano, L., Pichardo, M. C., & García-Berbén, A. B. (2017). Validation of the Spanish short self-regulation questionnaire (SSSRQ) through Rasch analysis. *Frontiers in psychology*, *8*, 276.

Großschedl, J., Mahler, D., & Harms, U. (2018). Construction and evaluation of an instrument to measure content knowledge in biology: The CK-IBI. *Education Sciences*, *8*(3), 145.

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in pharmacy teaching and learning*, *7*(6), 836-850.

He, P., Liu, X., Zheng, C., & Jia, M. (2016). Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. *Chemistry Education Research and Practice*, *17*(2), 381-393.

Herrmann-Abell, C. F., Hardcastle, J., & DeBoer, G. E. (2018). Using Rasch to Develop and Validate an Assessment of Students' Progress on the Energy Concept. *Grantee Submission*.

Kudiya, K., Sumintono, B., Sabana, S., & Sachari, A. (2018). Batik Artisans' Judgment of Batik Wax Quality and Its Criteria: An Application of the Many-Facets Rasch Model. In *Pacific Rim Objective Measurement Symposium (PROMS) 2016 Conference Proceedings* (pp. 27-37). Springer, Singapore.

Lambri, A., Mahamod, Z., Zakaria, N., Baharum, H., & Hanapi, Z. (2019). Validity and Reliability of the Malay Social Traits Instrument Using Rasch Measurement Model. *International Journal of Academic Research in Business and Social Sciences*, *9*(7), 225–232.

Lew, M. D., & Schmidt, H. G. (2011). Self-reflection and academic performance: is there a relationship?. *Advances in Health Sciences Education*, *16*(4), 529.

Linacre, J. M. (2012). A user's guide to WINSTEPS MINISTEP. *Rasch model computer programs. Beaverton, Oregon: Winsteps. com*.

Maat, S. M., & Rosli, M. K. (2016). The Rasch model analysis for statistical anxiety rating scale (STARS). *Creative Education*, *7*(18), 2820.

Maseko, J., Luneta, K., & Long, C. (2019). Towards validation of a rational number instrument: An application of Rasch measurement theory. *Pythagoras*, *40*(1), 441.

Mayes, R., Rittschof, K., Dauer, J., & Gallant, B. (2019). Quantitative modelling biology undergraduate assessment. *Letters in Biomathematics*, *6*(1), 1-27.

Misbah, M., Dewantara, D., Hasan, S. M., & Annur, S. (2018). The development of student worksheet by using Guided Inquiry Learning Model to train student's scientific attitude. *Unnes Science Education Journal*, *7*(1).

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the validity and reliability of research instruments. *Procedia-Social and Behavioral Sciences*, *204*, 164-171.

Park, M., & Liu, X. (2019). An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. *Research in Science Education*, 1-18.

Peeters, M. J., & Martin, B. A. (2017). Validation of learning assessments: A primer. *Currents in Pharmacy Teaching and Learning*, *9*(5), 925-933.

Pichardo, M. C., Cano, F., Garzón-Umerenkova, A., de la Fuente, J., Peralta-Sánchez, F. J., & Amate-Romera, J. (2018). Self-regulation questionnaire (SRQ) in Spanish adolescents: factor structure and rasch analysis. *Frontiers in psychology*, *9*, 1370.

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 020111.

Pleasence, P., & Balmer, N. J. (2019). Development of a general legal confidence scale: A first implementation of the Rasch measurement model in empirical legal studies. *Journal of Empirical Legal Studies*, *16*(1), 143-174.

Sabudin, S., Mansor, A. N., Meerah, S. M., & Muhammad, A. (2018). Validity and Reliability of Students' Science and Technology Culture Instrument (BST-M) using Rasch Measurement Model. *International Journal of Academic Research in Business and Social Sciences*, *8*(5), 986-995.

Susongko, P. (2016). Validation of science achievement test with the rasch model. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 268-277.

Upegui-Arango, L. D., Forkmann, T., Nielsen, T., Hallensleben, N., Glaesmer, H., Spangenberg, L., ... & Boecker, M. (2020). Psychometric evaluation of the Interpersonal Needs Questionnaire (INQ) using item analysis according to the Rasch model. *Plos one*, *15*(8), e0232030.

Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education*, *58*(8), 922-933.

Widhiarso, W., & Sumintono, B. (2016). Examining response aberrance as a cause of outliers in statistical analysis. *Personality and Individual Differences*, *98*, 11-15.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *45*(2), 224-238.

Yan, Z., Brown, G. T., Lee, J. C. K., & Qiu, X. L. (2020a). Student self-assessment: Why do they do it?. *Educational Psychology*, *40*(4), 509-532.

Yan, Z., Chiu, M. M., & Ko, P. Y. (2020b). Effects of self-assessment diaries on academic achievement, self-regulation, and motivation. *Assessment in Education: Principles, Policy & Practice*, *27*(5), 562-583.

Yang, Y., He, P., & Liu, X. (2018). Validation of an instrument for measuring students' understanding of interdisciplinary science in grades 4-8 over multiple semesters: A Rasch measurement study. *International Journal of Science and Mathematics Education*, *16*(4), 639-654.

Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and reliability learning transfer item using rasch measurement model. *Procedia-Social and Behavioral Sciences*, *204*, 212-217.

Yuhanna, W. L., & Retno, R. S. (2016). The learning of science basic concept by using scientifiq inquiry to improve student's thinking, working, and scientific attitude abilities. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *2*(1), 1-9.