JPII 10 (2) (2021) 241-252

Jurnal Pendidikan IPA Indonesia

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

DIAGNOSTIC TEST TO ASSESS MISCONCEPTIONS ON PHOTOSYNTHESIS AND PLANT RESPIRATION: IS IT VALID AND RELIABLE?

H. L. Lim*[1] and Y. P. Poo[2]

[1]School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia
[2]SMJK Phor Tay, Penang, Malaysia

DOI: 10.15294/jpii.v10i2.26944

Accepted: November 2nd 2020. Approved: June 29th 2021. Published: June 30th 2021

ABSTRACT

High percentage of secondary school students was found that they were lack of understanding of the relationship between photosynthesis and respiration in plants. They did not fully understand the importance and function about plant respiration. Thus, this study designed to develop a valid and reliable instrument in two- tier multiple choice questions format which called Photosynthesis and Plant Respiration Diagnostic Test (PRDT) to assess the common types of misconceptions related to this topic among form four students (Grade 10) in Malaysia. Survey research method was applied in this study. There were 500 participants from 15 secondary schools were involved. 45 minutes were given to the participants in answering 18 two- tier diagnostic test items. The psychometric properties of the instrument had been tested using Rasch analysis. The result found that the newly developed instrument was valid and reliable. It brought the significant contribution in teaching and learning, especially classroom assessment practice in Biology subject.

© 2021 Science Education Study Program FMIPA UNNES Semarang

Keywords: misconception; photosynthesis; plant respiration; two-tier multiple choice questions test

INTRODUCTION

Photosynthesis and respiration are important scientific topics in science curriculum of many countries, under the theme investigating the physiology of living things, students generally need to study about photosynthesis and plant respiration. Throughout the years, misconceptions about photosynthesis and plant respiration were well documented (Treagust & Haslam, 1986; Cullinane & Liston, 2011; Näs, 2012; Kestler, 2014; Galvin et al., 2015; Södervik et al. 2015; Anjarsari, 2018; Jayanti & Rahayu, 2019). For instance, in Australia, a high percentage of secondary school students was found that they were lack of understanding of the relationship between photosynthesis and respiration in plants.

***Correspondence Address**
E-mail: hllim@usm.my

They did not fully understand the importance and function about plant respiration. (Treagust & Haslam, 1986). On the other hand, a research conducted with the 13-year-old Greek students, revealed that photosynthesis was characterized by the large number of perspectives on the topic of energetical, ecological, physiological and biochemical. Thus it was found as a very difficult topic in biology (Marmaroti & Galanopoulou, 2006; Södervik et al., 2014). The findings of study revealed that most of the students have the incorrect conceptual understanding about photosynthesis and respiration although the teachers have taught it.

The problem of misconceptions in photosynthesis and plant respiration will prevent students' meaningful learning and permanent learning, it brought serious impact on students' future learning. The occurrence of these problem can hinder students from learning more advanced concepts related to this topic. It may result the students fail to grasp the new concepts or just learn for the test but revert to the misconceptions outside the classroom. Thus, the initial understanding of students must be carefully considered by teachers. One of the way is to identify the misconceptions that students already have from time to time, so the teachers can plan the more effective lessons in the teaching and learning process.

There are various assessment methods can be applied to identify students' misconceptions such as paper-and-pencil test, interview, concept map, drawing and word association, experiment. One of the most convenient method used is paper-and-pencil test. It can be administered to a group of students and the finding can be gained within a short period. Multiple-choice item test is the common format that specifically designed to detect students' various misconceptions and misunderstanding in a focus and limit content area (Treagust & Haslam, 1986; Treagust, 2012; Vila & Sanz, 2012).

The current Biology assessment tasks in paper-and-pencil test generally consists of closed questions that required only one word or one sentence of answers or districted-response essay questions test which allow students to develop their answers and demonstrate their understanding of a topic (Parker et al., 2012; Svandova, 2014; Gurel et al., 2015; Galvin et al., 2015; Akçay, 2016; Susanti, 2018). However, there are less systematic and practical way of diagnosing and analysing students' misconceptions. These limitations can be solved by using two- tier diagnostic test, as this type of test questions required students to give reasons to their answer. This will help to identify students' misconceptions in a clearer way, and how they do reasoning will reflect the level of their conceptual understanding. Based on Yang et al. (2017), two-tier multiple-choice diagnostic test serves to test the students' knowledge of facts and their reasoning skill, especially their justification of the answers in the first-tier (tier-1) (Treagust, 1988; Fulmer et al., 2015; Lin et al., 2016). It means that the newly developed diagnostic tests in the form of two-stage options was not only able to identify the types of misconceptions faced by students, but can provide deeper understanding of the student's cognitive thinking through the choice of student answers in the second-stage diagnostic test. Furthermore, this format was expected to gain more detail information of students' misconception problems that are categorized as correct answer at the first stage but wrong answer at the second stage. This situation cannot be identified with normal cognitive diagnostic test. Hence, the development of a valid and reliable two-tier multiple choice diagnostic test specifically for the purpose of identifying students' misconceptions about photosynthesis and plant respiration was required.

The developed diagnostic test helped to identify which learning objectives of photosynthesis and plant respiration that the students were unable to master and the possible causes why they were unable to master the particular learning objectives (Nitko & Brookhart, 2014). Thus, it became a powerful and useful assessment tool to help the teachers to detect their students' misconception problems that occur during the teaching and learning process (Treagust & Haslam, 1986; Treagust, 2012). Therefore, this study aimed to develop and validate a two- tier multiple choice diagnostic test on the area of photosynthesis and plant respiration. The descriptive survey research was carried out using quantitative approach, with 500 participants. The reliability and validity of newly developed PRDT instrument was tested with the Rasch Model.

Various format of diagnostic tools had been developed to investigate the misconceptions about photosynthesis and plant respiration. Köse (2008) claimed that the drawing method in combination with interviews method had become a useful method to diagnose the student's conceptual understandings and misconceptions about the abstract concepts, such as the topic photosynthesis and plant respiration. For instance, a research was conducted in Turkey, which data gathered from the drawings of 156 students aged 20-25 from four classes and the interview of 15 students. The findings of this researched was found matched with other studies in this field.

Parker et al. (2012) presented the findings related to the undergraduates' thinking about photosynthesis using Diagnostic Question Cluster (DQC). The data collected from various item formats had been triangulated and compared, namely students' responses to the cluster items, multiple-choice items, multiple-true/false items and essay items. The data collected from multiple-true/false items revealed that the students faced a mixture of accurate and inaccurate ideas. Besides, the interview data about the students' choice on the multiple-choice item revealed the insubstantiality of students' understanding on this topic. The data exhibited that many undergraduates lack of ability in the: (i) basic understanding about the role of photosynthesis in plant metabolism; and (ii) reasoning with the scientific principles when learning the new contact.

Meanwhile, Gurel et al. (2015) presented an overview of the common diagnostic instruments used in assessing students' misconceptions in science. The findings showed that interview was the most common method used (53%), followed by open-ended item (32%), and multiple-choice item (13%). On the other hand, Galvin et al. (2015) applied a diagnostic test consisted of 19 multiple-choice items to identify the misconceptions in respiration and photosynthesis. The scenario of items was presented in text and picture, addressing the key research questions. There were five options for each item. One of the option was the key answer, one was the misconception identified in the literature, along with other two distractors and an option to be chosen if they did not know the answer. The diagnostic test was found beneficial to a pedagogical cycle for the recognition, reduction and removal of misconceptions.

Based on the findings of these studies, it can be concluded that each assessment method and assessment tool has the advantages as well as the disadvantages that should be kept in mind about their usages. The teachers and researchers must be aware of the selection for the most effective and appropriate method and instrument in order to identify the students' misconceptions in the most effective way. In this study, the researchers had developed a two-tier diagnostic test to identify students' misconceptions in a clearer way, and how they do reasoning will reflect the level of their conceptual understanding.

Treagust (1986) and Treagust (2012) suggested ten stages involved in three broad areas on the development of a two-tier diagnostic tests for identifying students' misconceptions in photosynthesis and plant respiration. According to the author, the first area in the development of diagnostic test was to define the content, involved four stages. Researcher needed to first identify the propositional knowledge statements related to the area of study, followed by the development of a concept map and relates propositional knowledge to the concept map. Lastly, all propositional statements needed to be validated by professional in the area of study to ensure the content validity. The second area in test development was to obtain information about students' misconceptions. This area was fulfilled by examining related literatures, conducting unstructured student interviews and developing multiple choice content items with free response to find out students' misconceptions in this area. All information obtained in this area will be the guidance in the development of two-tier diagnostic test. The last area in the development of diagnostic test was the development of the real diagnostic instrument, involved three main stages. The development of two-tier diagnostic test would be followed by designing a specification grid to show the propositional statements addressed by each of the items in the test. Lastly, the instrument developed needed to be further revise for refinement. All these steps were applied to develop the two- tier diagnostic test of this study.

METHODS

Descriptive survey design was used in this study. The function of a survey was to collect original data for describing a population too large to observe directly (Lowhorn, 2007). Furthermore, this research design was appropriate to be used to investigate human's behaviour, belief, attitude, idea or knowledge of a particular group or individual as it provided an accurate interpretation of human characteristics. Hence, this research design was selected to diagnose misconceptions in the topic photosynthesis and plant respiration among form four students.

The sample comprised of 500 form four (tenth grade) students from 15 secondary schools in Penang State, Malaysia. They were selected based on the recommendation of sample size and item calibration stability table (Linacre, 2012). For Rasch Model, the requirement of number of items and number of samples are symmetry to produce statistically stable measures for person and item (Linacre, 2012). In every school, sample included all students from middle performance classes and low performance classes. This is because students in high performance classes normally score well in examination, meaning their understanding of scientific concepts are better compared to those who are not performing well in examination.

The nine steps suggested by Treagust (1988) and Treagust (2012) was applied in the development of Plant Respiration Diagnostic Test (PRDT). First, step 1. Form 4 Biology Integrated Curriculum for Secondary Schools Curriculum Specifications was the main reference to develop all the items to ensure the alignment between the items relevance and learning objectives of the syllabus. Second, step 2 & 3. Concept map which related to photosynthesis and plant respiration was developed. The purpose was to carefully analysed and categorised the nature of the content domain selected for the assessment (Treagust, 1988; Treagust, 2012). Then, the internally consistent of the content domain was determined by relating directly the photosynthesis and plant respiration to the concept map (Treagust, 1988). Third, step 4. Five experienced science teachers who taught Form Four Biology were asked to check the content of the propositional statements in the concept map to ensure content validity, Experts were requested to score each propositional statement from 1 to 4, where 1 represented the statement is not relevant, 2 and 3 represented that the statement was useful but need revision, and 4 represented that statement was very relevant. Content validity ratio was calculated based on the formula CVR = (Ne - N/2) / (N/2), where Ne was the number of panels indicating "propositional statement is essential" and N wass the total number of panels. In this study, CVR 0.99 was needed to ensure the content validity, as the numbers of panel were five. The numeric value of content validity ratio was determined by Lawshe Table (1975). Based on the calculations, the value of CVR for all the propositional statements were reported at 1.00. This means all panels agreed that the propositional statements in the concept map were scientifically accurate and suitable to measure what they supposed to. Fourth, step 5 & 6. A pre test instrument contained 24 open ended questions developed by researcher was used to identify students' misconceptions in the topic photosynthesis and plant respiration. The content validity of the items was checked by 5 panels. Then, the CVR of every item in the pre test was calculated. In this study, CVR 0.99 is needed to ensure the content validity, as the numbers of panel were five. All the items with CVR values lesser than 0.99 were revised. Table 1 showed the item specification table for pre test.

Table 1. Item Specification Table for Pre Test

Domain	Key Ideas	Item
Photosynthesis	Anabolic process (the synthesis of carbohydrate).	1,11
	Substances required for photosynthesis are carbon dioxide and water.	1,3,13
	Substances produced from photosynthesis are glucose and oxygen. Energy is stored in carbohydrate molecules.	1,3,15
	Occurs only in cells containing chlorophyll.	5
	Occurs only in the presence of sunlight.	17
	Chloroplast is the site of reaction.	1,7,19 9,21,22
Plant Respiration	Catabolic process (the breakdown of carbohydrate).	2,12
	Substances required for respiration are glucose and oxygen.	2,4,14
	Substances produced from respiration are carbon dioxide and water.	2,4,16
	Energy is liberated in the form of ATP.	6
	Occurs in all plants at all times.	8,20
	Independent of chlorophyll and sunlight.	18
	Mitochondrion is the site of reaction.	10,23,24

Fifth, step 7. Items in the format of open-ended question were used in the pre test. This format allowed the respondents to supply more detail information in their responses, including their thinking process, ideas and their understandings of the topic. Sixth, step 8. The Photosynthesis and Plant Respiration Diagnostic Test (PRDT) was a two- tier diagnostic test consisted 18 items developed by the researcher. The test items were prepared based on learning objectives and outcomes, propositional statements, category of misconceptions (dimension) and from students' responses in the open-ended pre test items. Table 2 showed the distribution of items in PDRT based on different dimensions.

Table 2. Distribution of PRDT Items Based on Different Dimensions

Dimension (Misconception)	Item No.		Total Item
	Photosynthesis	Respiration in Plant	
Not tracing matter	1a,1b,2a,2b, 5a,5b,	3a,3b,4a,4b, 6a,6b	12
Not tracing energy	7a,7b, 9a,9b,11a,11b,	8a,8b, 10a,10b, 12a,12b	12
Not tracing scale and location	13a,13b, 15a,15b, 18a*,18b*	14a,14b,16a,16b, 17a,17b, 18a*,18b*	12

*covered both content photosynthesis and respiration in plant

There were two tiers or two parts in the format of two-tier diagnostic test item. The first part of the item, representing the content knowledge to be assessed, namely the content about photosynthesis and plant respiration. There were two choices given for this part, namely "Yes" and "No". Meanwhile, the second part consisted of the reasons for choosing the responses for the first part. There were four options provided as possible reasons for the responses to the first part. The reasons consisted of the designated correct responses together with the distractors which identified by open ended pre test questions, and some based on literature reviews. Seventh, step 9. Two experienced English teachers were asked to check the language of the instrument to make sure the language used was clear and not bias to any group of respondents. At the same time, 5 experienced Biology teachers were asked to check for the content validity of items in PRDT, followed the same scoring criteria and calculation of CVR values as done with pre test items.

Every student was given 40 minutes to complete the paper and pencil test. Instrument was personally distributed by the researcher. Data collected from the PRDT was analysed quantitatively using Rasch analysis. According to the dichotomous Rasch Model, the difference in the respondent's overall performance level (ability level) and the difficulty of the items determined the probability of the respondent to respond to an item correctly, the related equation was as follows:

$$ln\left(\frac{P_{ni}}{1-P_{ni}}\right) = B_n - D_i$$

P_{ni} is the probability that student n of overall performance level B_n will respond correctly to item i with a difficulty of D_i (Bond & Fox, 2007). In Rasch measurement, individuals' ability and items' difficulty were estimated simultaneously (Fulmer et al., 2015). They could be compared on a common scale, called a logit scale, based on the chance that each individual would answer each item correctly (Fulmer et al., 2015). WINSTEPS (Linacre, 2012) was used to estimate the students' understanding of the topic and the item difficulties. According to Rasch model, the student who have a greater ability than others should have the greater probability of solving the item correctly. In the other words, less difficult item, the greater probability to answer correctly compared to the more difficult item.

The degree of validity of assessment was determined based on Rasch analysis. First, the unidimensionality of the PRDT was tested to ensure the assessment was reinforced by a single attribute. Next, the others analysis to support the validity indices were also shown, namely item difficulty, item fit, item and person reliability and gender differential item functioning (DIF) (Yao & Mok, 2012). All this analysis aimed to confirm that the PRDT was appropriate, valid and reliable diagnostic assessment tool for diagnosing the misconceptions of students in the topic of photosynthesis and respiration in plants. The unidimensionality of the assessment was conducted and tested through a Principal Components Analysis of Rasch residuals subroutine in the Winsteps software (Linacre, 2012). It was recommended that an acceptable criterion for establishing the unidimensionality could be based on the eigenvalue of the first contrast. It should less than 2.0 (Linacre, 2012).

Item difficulty index of the items was analysed to find out the difficult and easy items in relation to the participants' abilities. Further information could be found by analysing the Wright Map. It was a visual representation that showed the distribution of the respondents' ability in relation to the various level of difficulty of the items (Yao & Mok, 2012). The right map showed the distribution of respondents' abilities on the left side of a vertical line and the difficulty of the items on the right side of the line (Herrmann & George, 2011). Meanwhile the item and person separation and reliability indicated the items spread across the trait continuum (Linacre, 2012). Higher separation index yields higher reliability.

Differential Item Functioning (DIF) considered the important issue in establishing assessment fairness. It detected the probabilities of getting correctly to an item when the respondents have the similar ability level but from different backgrounds, such as gender, location, social-economic status. According to Yao and Mok (2012), the magnitude of DIF indicated the extent to which the item parameter differs from different groups.

RESULTS AND DISCUSSION

Table 3 showed the reliability index and separation index of person and item in this study. Analysis was performed on the PRDT instrument as a whole, namely the reliability and the separation of the item and the person.

Table 3. Reliability Index and Separation Index

	Reliability Index	Separation Index
Person	0.79	1.92
Item	0.99	8.47

Based on Table 3, the reliability of item was 0.99, while the separation of item was 8.47. The reliability of the items 0.99 was in very good condition, as according to Fisher (2007), reliability more than 0.94 was considered very good. High item reliability meaning that PRDT has acceptable number of items to measure what was supposed to be measured under the topics photosynthesis and respiration in plant. Besides, the separation index of item was 8.47, when rounded off was equal to 8.0 indicated the items could be separated into 8 groups according to the responses by students. The separation of items showed a good value as according to Linacre (2012), which described the separation of more than 2.0 was a good value.

On the other hand, the person reliability showed a value of 0.79 while the person separation index was 1.92. The person reliability 0.79 was considered good and acceptable, as according to Fisher (2007), reliability more than 0.70 was good. The person separation index was 1.92, when rounded off was equal to 2.0 indicated the person could be separated into almost 2 groups by the items of PDRT. This value was still acceptable as according to Linacre (2012), which described the separation of more than 2.0 was good. The lower person separation index means there was less information available to estimate the student measures, which resulted in lower person reliability, 0.79. In contrast, high item separation index and high item reliability index were due to large sample size, n was 500 in this study, where the differences in difficultly level of the items were easier to determine. In other words, it was easy to separate 36 items by 500 students, but it was comparatively more difficult to separate 500 students by only 36 items.

Construct validity was a degree in which a test measured what it was supposed to measure. In this study, researcher determined the construct validity of PRDT by focusing on the analysis of unidimensionality, fit statistic, item map and differential item functioning (DIF). Table 4 showed the standardized residual variance of PDRT. Unidimensionality was an important aspect to ensure the measurement of PRDT was specific to one construct. In order to assess unidimensionality, Principal Component Analysis (PCA) of the Rasch residuals was performed to find out how much variance of PRDT measuring what it was supposed to measure.

Table 4. Table of Standardized Residual Variance (in Eigenvalue units)

	Empirical		Modeled
Total variance in observations	65.8	100.0%	100.0%
Variance explained by measures	29.8	45.3%	44.3%
Unexplained variance (total)	36.0	54.7%	55.7%
Unexplained variance explained by 1st factor	2.7	4.1%	
Unexplained variance explained by 2nd factor	2.5	3.7%	
Unexplained variance explained by 3rd factor	2.2	3.3%	
Unexplained variance explained by 4th factor	1.9	2.9%	
Unexplained variance explained by 5th factor	1.8	2.7%	

Results were shown in Table 4. The raw variance explained by measures was 29.8 unit, comprised of 45.3% which was closely matched the expected 44.3%. The Rasch cut off point of 40% is achieved, validated a unidimensional trait of the data. The findings showed that all the items in PRDT fit the model well. It indicated that the wide spreading of item and persons as well as the relatively good item person targeting. Besides, there were small amounts of unexplained variances in the components which came from the residuals, 4.1%, 3.7%, 3.3%, 2.9% and 2.7% for the first, second, third, fourth and fifth factors respectively. Eigen value for unexplained variance explained by 1st factor was 2.7 units, represented 4.1% of the residual variance, and indicated that the biggest factor from the residual was only 3 items.

Table 5 and Table 6 showed the summary of fit model and mean measure of PRDT for 36 items and 497 respondents (non- extreme scores).

Table 5. Summary of Measured 500 Persons

	Raw Score	Count	Measure	Model Error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	23.0	36.0	0.80	0.41	1.00	0	0.98	0
S.D.	5.7	0	0.94	0.09	0.16	1.0	0.31	1.0
Max.	35.0	36.0	3.93	1.02	1.72	4.0	3.05	4.1
Min.	10.0	36.0	-1.12	0.67	0.67	-2.9	0.40	-2.0

From Rasch analysis, person mean gave a value of 0.80 logit, indicated that person met the expectation. Item mean from Table 6 was set to an arbitrary 0.00. Zero was setting all items to give by definition a 0.5 probability of a correct response in the Rasch model.

Table 6. Summary of Measured 36 Items

	Raw Score	Count	Measure	Model Error	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	317.2	497.0	0	0.11	1.00	0.1	0.98	0
S.D.	81.4	0	0.99	0.02	0.08	2.2	0.15	2.1
Max.	467.0	497.0	1.76	0.19	1.19	5.4	1.28	5.4
Min.	151.0	497.0	-2.27	0.10	0.84	-5.0	0.50	-4.3

The values of mean infit MNSQ for both person and item were 1.00, same as expected by model. The values of outfit MNSQ for both person and item were 0.98, indicated that the model predicted the data too well. In term of MNSQ, the range 0.50 to 1.50 supported productive measurement (Linacre, 2012). The values of mean infit ZSTD for person and item were 0.00 and 0.10 respectively, the mean values of outfit ZSTD were same, reported 0.00. The values of 0.00 for ZSTD were same as expected by model, while the value of ZSTD was more than 0.00 indicated lack of predictability. Item with value of MNSQ nearer to 1.00 and value of ZSTD nearer to 0.00 was deemed a better fit (Linacre, 2012). As a conclusion, all data reported to be overall fit and accepted by the Rasch model. Table 7 showed item statistic with misfit order.

The finding shown that the mean value for infit MNSQ was 1.00, with standard deviation 0.08, while the mean value for outfit MNSQ was 0.98 with standard deviation 0.15. The range of fit in this study followed the range 0.7 to 1.3 for multiple choice tests (Bond & Fox, 2007). The infit values of MNSQ obtained were in between 0.84 to 1.19. The outfit values of MNSQ were in the range of 0.79 to 1.28. Results showed all items of PRDT were fit under the model. The values of PTMEA CORR shown all were positive, with the range 0.20 to 0.54. The positive values of PTMEA CORR showed that all items were carefully constructed (Bond & Fox, 2007), where all items were able to distinguish between the ability of respondents. In Rasch analysis, the predictability of items was examined directly using the MNSQ statistic, rather than indirectly through the correlations (Linacre, 2012).

Table 7. Item Statistic with Misfit Order

Item	Raw Score	Count	Measure	Model S.E.	Infit		Outfit		PTMEA CORR.
					MNSQ	ZSTD	MNSQ	ZSTD	
9a	251	497	0.76	0.10	1.19	5.4	1.28	5.4	0.20
13a	338	497	-0.10	0.10	1.03	0.8	1.22	2.7	0.30
9b	218	497	1.07	0.10	1.14	3.8	1.20	3.8	0.26
12a	347	497	-0.20	0.10	1.08	1.8	1.15	1.9	0.26
2b	260	497	0.67	0.10	1.11	3.1	1.15	2.9	0.28
7b	233	497	0.93	0.10	1.14	3.9	1.14	2.8	0.27
7a	207	497	1.18	0.10	1.09	2.4	1.09	1.8	0.32
8a	238	497	0.88	0.10	1.07	2.0	1.09	1.9	0.32
17a	303	497	0.26	0.10	1.08	2.1	1.06	1.1	0.30
11b	398	497	-0.82	0.12	1.00	0.1	1.07	0.6	0.27
18b	151	497	1.76	0.11	1.03	0.6	1.06	0.9	0.37
12b	302	497	0.27	0.10	1.02	0.6	1.05	0.9	0.34
5b	308	497	0.21	0.10	1.05	1.3	1.04	0.7	0.32
16a	332	497	-0.04	0.10	1.05	1.2	1.04	0.6	0.31
5a	405	497	-0.92	0.12	1.03	0.5	1.04	0.4	0.25
10a	345	497	-0.17	0.10	1.01	0.3	1.03	0.4	0.33
1a	448	497	-1.71	0.16	1.02	0.2	0.97	-0.1	0.21
18a	328	497	0.01	0.10	1.01	0.3	0.96	-0.6	0.35
1b	394	497	-0.76	0.12	1.00	0.0	0.97	-0.2	0.30
15a	460	497	-2.04	0.18	0.99	0.0	0.91	-0.3	0.20
2a	461	497	-2.07	0.18	0.98	-0.1	0.81	-0.8	0.23
16b	310	497	0.19	0.10	0.98	-0.5	0.98	-0.3	0.38
11a	295	497	0.34	0.10	0.98	-0.6	0.95	-0.9	0.40
8b	201	497	1.24	0.10	0.95	-1.3	0.97	-0.6	0.44
3a	321	497	0.08	0.10	0.97	-0.8	0.92	-1.3	0.39
6a	240	497	0.86	0.10	0.96	-1.1	0.95	-1.0	0.43
10b	244	497	0.72	0.10	0.93	-2.1	0.90	-2.1	0.45
13b	362	497	-0.36	0.11	0.93	-1.5	0.88	-1.4	0.40
15b	429	497	-1.31	0.14	0.92	-0.9	0.81	-1.2	0.34
6b	221	497	1.04	0.10	0.91	-2.5	0.90	-2.2	0.48
14a	467	497	-2.27	0.19	0.91	-0.5	0.50	-2.2	0.30
4a	297	497	0.32	0.10	0.91	-2.6	0.84	-2.9	0.47
3b	317	497	0.12	0.10	0.89	-2.9	0.84	-2.5	0.46
14b	410	497	-1.00	0.12	0.89	-1.5	0.79	-1.7	0.38
17b	334	497	-0.06	0.10	0.84	-4.0	0.77	-3.4	0.50
4b	233	497	0.93	0.10	0.84	-5.0	0.81	-4.3	0.54
MEAN	317.2	497	0.00	0.11	1.00	0.1	0.98	0.0	
S.D.	81.4	0.00	0.99	0.02	0.08	2.2	0.15	2.1	

In the Rasch analysis, the item- person map was the visual representation of the distribution of the students' abilities in relation to the distribution of the item difficulties. It provided a clear picture of the multiple choices PDRT items by placing the difficulty of all items on the same measurement scale as the abilities of students. From the item- person map, researcher had a better understanding on how well the PDRT measured.

Figure 1 showed the item- person map for the entire set of 18 items (every items consist of two tiers, labelled a and b) in PRDT. The left hand side of the map showed the range of students' performance, while the right hand side of the map showed item's difficultly. Low performance of student and low difficulty of item represented at the bottom of the map and vice versa. The mean of the item difficultly was set to be zero, so that students had a 50% of chance to answer the item correctly.

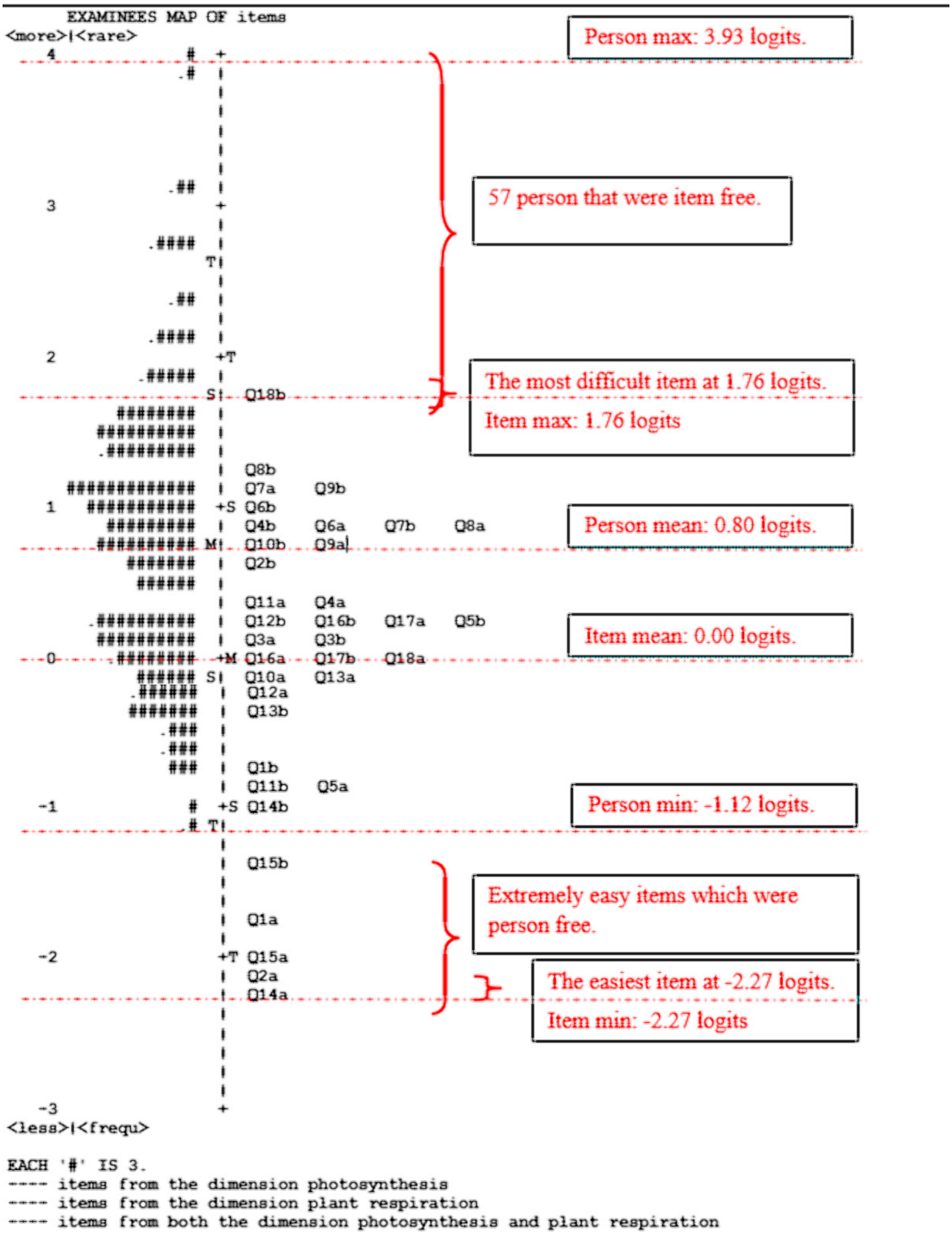

Figure 1. Item- Person Map for the 18 Items

According to Figure 1, the range of items' difficulty of PRDT was about 4 logits, in the range from -2.27 logits to 1.76 logits. Item 18b at the 1.76 logits was the most difficult item. Item 18b was aimed to diagnose students' misconception that related to trace scale and location in photosynthesis and respiration in plant. Many students failed to choose the correct response that all living cells respire and cells with chloroplast could carry out photosynthesis, they had misconception that photosynthesis and respiration only happened in leaves. On the other hand, item 14a at -2.27 logits was reported as the easiest item. Item 14a was aimed to detect whether students understood that all plants carry out respiration. All students answered correctly. Besides, according to the item-person map, there were five extremely easy items (15b, 1a, 15a, 2a and 14a) which were person free. From these 5 extremely easy items, four of them (1a, 15a, 2a and 14a) were items in the first tier of PDRT, which were designed to detect students' misconceptions on content level, whereas item 15b was in the second tier of PDRT, designed to assess students' misconception on reasoning level. Last but not least, items in PRDT were separated in six gaps, which could be noticed between item 15a and 1a, 1a and 15b, 15b and 14b, 1b and 13b, 11a/4a and 2b, 8b and 18b from the item-person map.

Students' performance showed a range from -1.12 logits to 3.93 logits, which was about 5 logits. From the item- person map, 57 students were in the category of item free. These 57 students performed better as compared to the most difficult item which was only at 1.76 logits. These students required more difficult task. As a conclusion, the difficulty levels of items in PRDT were considered easy in the overall, as the distributions of majority of the items were below the person mean value, which was 0.80 logits.

Table 8 showed DIF cases with corresponding t- value. All items in PRDT must not behave differently for particular subgroups of students to ensure a higher degree of construct validity. The Rasch measurement model enabled the detection of items which were biased toward different subgroups according to construct irrelevant factors such as ability, gender and ethnicity of students. In this study, the differential item functioning (DIF) was studied. Items with t- value less than -2.0 and more than 2.0 were considered DIF items, and were further suggested to be revised or to be eliminated (Alquraan et al., 2010).

Table 8. DIF Items with Corresponding t- Value

DIF item	t- value
1a	-4.82
1b	-7.18
3a	2.26
5a	-2.16
6a	2.52
6b	3.15
7a	3.31
8b	2.53

There were total 8 DIF cases detected in this instrument, tier one and tier two of item 1 (la and 1b), tier one of item 3 (3a), tier one of item 5 (5a), tier one and tier two of item 6 (6a and 6b), tier one of item 7 (7a) and tier one of item 8 (8a). Majority of the DIF cases were from tier one of the items. Item 1, item 5 and item 7 were in the domain photosynthesis, whereas item 3, item 6 and item 8 were from the domain plant respiration. All the DIF items will be further revised. Apart from that, 28 cases showed t- values in the range of $-1.37 < t < 1.99$, which indicated that these items are not biased and can be used in measuring the construct (Alquraan et al., 2010).

The positive value of PTMEA CORR showed that the item was able to contribute to the assessment of the competitiveness scale. In other words, the item was able to distinguish between the ability of the students. In this study, the PTMEA CORR values were positive for all the items. According to Jingjing and Magdalena (2013), Bond and Fox (2007), it shown that the items were able to measure the construct to be measured. It contributes to the good degree of construct validity.

Besides, the finding revealed that students'' performance showed a range from -1.12 logits to 3.93 logits, From the item- person map, there were 57 students in the category of item free. They performed better as compared to the most difficult item which was only at 1.76 logits. As a conclusion, the items in PRDT were considered easy in the overall, as the distributions of majority of the items were below the person mean value. However, it was expected as the PRDT was not measuring the students' mastery of the content but to diagnose students' misconceptions. Thus, majority of the high ability students were definitely able to respond the difficult item correctly (Chang & Lo, 2015; Herrmann & DeBoer, 2016).

In comparison with previous study by Treagust & Haslam (1986), Svandova (2014), Wind dan Gale (2015), and Herrmann and DeBoer (2016) in the diagnostic test, which reported a value of 0.72, 0.43, 0.66 and 0.98 for the instrument reliability, PRDT was showing a better result, reported at 0.99. The same conclusion was made when comparing the index of item reliability of PRDT with the reliability index reported from a research carried out by Kılıç and Sağlam (2009) in the area of genetics concepts, which was only 0.86. High item reliability index in this research was the results of large item difficulty range and a large sample of persons (Linacre, 2012).

Another difference of this research with previous studies was the use of Rasch data analysis. In the previous studies, classic analysis approach was applied which contains two limitations, namely item dependent and person dependent. The person's observed score always depended on the difficulty of item. A person would have higher score if the items were easy and lower score when the items were difficult. This means, the response of the items was affected by the person involved in the test (Parker et al., 2012; Svandova, 2014; Gurel et al., 2015; Galvin et al., 2015; Akcay, 2017; Susanti, 2018). Meanwhile, the use of Rasch analysis enable the estimation of reliability and validity regardless the persons and their underlying cognitive ability because it enables the estimation of the item difficulty and person's ability along an interval scale that used a unit of standard measurement which called logit scale (Teh & Lim, 2016; Goh et al., 2017). It was a psychometric technique that was developed to improve the precision with which researchers construct instruments, monitor instrument quality, and compute respondents' performances (Boone, 2016). Compared to the traditional method of data analysis, Rasch analysis helped researchers to think in more sophisticated ways with respect to the constructs they wished to measure. Furthermore, the use of option probability curves to analyse how students responded to every items were new in the topic of photosynthesis and plant respiration.

In this study, there were eight Differential Item Functioning (DIF) cases reported. DIF occurred when people with the same ability level but from different groups showed differences in the probability of answering correctly. DIF have been increasingly applied in bias analysis with the rising concerns over the fairness of test. A test is considered bias if the factor that brought such a difference is not part of the construct or the focus in the test. The focus of this DIF analysis in this study was gender. However, the decision as to whether the real source of DIF in an item is part of the bias is totally subjective (Pae, 2011). In this study, 3 experts who were experienced Biology teachers were consulted to discuss about the DIF cases reported. Experts suggested that all the eight DIF items were appropriate to be used to detect misconceptions among students.

CONCLUSION

This research would bring more quality items to question bank. Items that have high degree of validity and reliability in detecting students' misconceptions were good items to be used by teachers to investigate students' understanding. School teachers could easily access questions regarding these topics and used them to create a more quality assessment. Besides, teachers might select some items from question bank to conduct remedial classes to students who were weak in understanding the concept of photosynthesis and respiration in plant too. The diagnostic of students' misconceptions on photosynthesis and respiration required the valid and reliable assessment tool. Hence, it was vital to conduct more researches related to this context in order to inform the teachers about the appropriateness of the assessment tool. Although there were a number of similar assessment tools had been developed by others countries, but it must be carefully adapted or referred due to the factors of differences in curriculum, education system, language and culture.

REFERENCES

Akçay, S. (2016). Prospective elementary science teachers' understanding of photosynthesis and cellular respiration in the context of multiple biological levels as nested systems. *Journal of Biology Education, 51*(1). 52-65.

Alquraan, M., Alshraideh, M., & Bsharah, M. (2010). Psychometric properties and differential item functioning (DIF) analyses of Jordanian version of Self-Assessed Wisdom Scale (SAWS-Jo). *International Journal of Applied Educational Studies*, *9*(1), 52-66.

Anjarsari, P. (2018). The Common Science Misconceptions in Indonesia Junior High School Students. *Journal of Science Education Research*, *2*(1), 21-24.

Bond, T. G., Fox, C. M., & Lacey, H. (2007). Applying the Rasch model: Fundamental measurement. In *in the social sciences 2nd*.

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

Chang, L., & Lo, S. J. (2015). A Rasch Model Distractor Analysis on Ordered Multiple-Choice Items of a Fractional Test. *Jiaoyu Yu Xinli Yanjiu= Journal of Education & Psychology*, *38*(2), 87.

Cullinane, A., & Liston, M. (2011). Two-tier Multiple Choice Questions (MCQs)-How Effective are They: A Pre-service Teachers Perspective. *IOSTE-NW Europe*, 1-12.

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095.

Fulmer, S. M., D'Mello, S. K., Strain, A., & Graesser, A. C. (2015). Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemporary Educational Psychology*, *41*, 98-110.

Fulmer, G. W., Chu, H. E., Treagust, D. F., & Neumann, K. (2015). Is it harder to know or to reason? Analyzing two-tier science assessment items using the Rasch measurement model. *Asia-Pacific Science Education*, *1*(1), 1-16.

Galvin, E., Simmie, G. M., & O'Grady, A. (2015). Identification of misconceptions in the teaching of biology: A pedagogical cycle of recognition, reduction and removal. *Higher Education of Social Science*, *8*(2), 1-8.

Goh, H. E., Marais, I., & Ireland, M. J. (2017). A Rasch model analysis of the mindful attention awareness scale. *Assessment*, *24*(3), 387-398.

Gurel, D. K., Eryılmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science.

Herrmann-Abell, C. F., & DeBoer, G. E. (2016). Using Rasch Modeling and Option Probability Curves to Diagnose Students' Misconceptions. *Grantee Submission*.

Jayanti, P. (2020). Comparative Study: Misconceptions On Photosynthesis And Respiration Concepts From Past To The Present. *JPPS (Jurnal Penelitian Pendidikan Sains)*, *9*(1), 1750-1755.

Kestler, N. S. (2014). *Middle School Science Teachers' understanding of students' misconceptions of photosynthesis and respiration* (Doctoral dissertation).

Kılıç, D., & Sağlam, N. (2009). Development of a two-tier diagnostic test concerning genetics concepts: the study of validity and reliability. *Procedia-Social and Behavioral Sciences*, *1*(1), 2685-2686.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology*, *28*(4), 563-575.

Lin, D., Talathi, S., & Annapureddy, S. (2016, June). Fixed point quantization of deep convolutional networks. In *International conference on machine learning* (pp. 2849-2858). PMLR.

Linacre, J. M. (2012). A User's Guide to Winstep. Ministep Rasch-Model Computer Programs.: Program Manual 3.73. 0. 2011.

Lowhorn, G. L. (2007, May). Qualitative and quantitative research: How to choose the best design. In *Academic Business World International Conference. Nashville, Tennessee*.

Marmaroti, P., & Galanopoulou, D. (2006). Pupils' understanding of photosynthesis: A questionnaire for the simultaneous assessment of all aspects. *International Journal of Science Education*, *28*(4), 383-403.

Näs, H. (2012). Understanding Photosynthesis and Respiration: Is It a Problem? Eighth Graders' Written and Oral Reasoning About Photosynthesis and Respiration. In *Issues and challenges in science education research* (pp. 73-91). Springer, Dordrecht.

Nitko, A. J.,& Brookhart, S. M. (2014). *Educational Assessment of Students (6th Edition)*. USA: Pearson.

Pae, H. (2011). Research Note: Differential Item Functioning and Unidimensionality in the Pearson Test of English Academic.

Parker, J. M., Anderson, C. W., Heidemann, M., Merrill, J., Merritt, B., Richmond, G., & Urban-Lurain, M. (2012). Exploring undergraduates' understanding of photosynthesis using diagnostic question clusters. *CBE—Life Sciences Education*, *11*(1), 47-57.

Södervik, I., Virtanen, V., & Mikkilä-Erdmann, M. (2015). Challenges in understanding photosynthesis in a university introductory biosciences class. *International Journal of Science and Mathematics Education*, *13*(4), 733-750.

Susanti, R. (2018, May). Misconception of biology education student of teacher training and education of Sriwijaya University to the concept of photosynthesis and respiration. In *Journal of Physics: conference series* (Vol. 1022, No. 1, p. 012056). IOP Publishing.

Svandova, K. (2014). Secondary school students' misconceptions about photosynthesis and plant respiration: Preliminary results. *Eurasia Journal of Mathematics, Science and Technology Education*, *10*(1), 59-67.

Teh, J., & Lim, H. L. (2016). Examining Psychometric Properties of Malay Version Children Depression Invento-ry (CDI) and Prevalence of Depression among Secondary School Students. *Pertanika Journal of Social Sciences & Humanities*, *24*(4).

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International journal of science education*, *10*(2), 159-169.

Treagust, D. F. (2012, October). Diagnostic assessment in science as a means to improving teaching, learning and retention. In *Proceedings of The Australian Conference on Science and Mathematics Education (formerly UniServe Science Conference)*.

Treagust, D. F., & Haslam, F. (1986). Evaluating Secondary Students' Misconceptions of Photosynthesis and Respiration in Plants Using a Two-Tier Diagnostic Instrument.

Vila, F., & Sanz, A. (2012). Reintroducing photosynthesis. *Biochemistry and Molecular Biology Education*, *40*(2), 148-155.

Wind, S. A., & Gale, J. D. (2015). Diagnostic opportunities using Rasch measurement in the context of a misconceptions-based physical science assessment. *Science Education*, *99*(4), 721-741..

Yao, J. J. J., & Mok, M. M. C. (2013). Implementing assessment for learning: An application of the Rasch model for the construction of a mathematics assessment to inform learning.