Three-tier Test to Identify Students' Misconception of Human Reproduction System

Wahid Hasyim[*], Hadi Suwono, Herawati Susilo
Biology Education–Universitas Negeri Malang
Jl. Semarang 5 Malang, Indonesia 65145. E-mail: hasyimwahid9@gmail.com[*]

Abstract: The purpose of this research was to develop a three-tier test of the human reproductive system in identifying student misconceptions. The instrument was developed by Treagust model which has three stages of defining content; obtain information; development of diagnostic tests. A total of 100 students were subjected. All question items consist of 15 items have been classified as valid, with result of reliability analysis equal to 0.717, average test of power difference 0.45, all item problem is in range 0.35–0.67 at level of difficulty, construct validity r = 0.531, N = 100, p = 0.000. False negative content validity test is 6.47% and false positive is 9.07%. Based on the analysis result, this instrument is feasible to identify student misconception.

Key Words: human reproductive system, misconception, three-tier test, student

Abstrak: Tujuan penelitian ini untuk mengembangkan instrumen *three-tier test* sistem reproduksi manusia dalam mengidentifikasi miskonsepsi siswa. Instrumen dikembangkan dengan model Treagust yang memiliki tiga tahapan yaitu mendefinisikan konten; memperoleh informasi; pengembangan tes diagnostik. Sebanyak 100 siswa dijadikan subjek. Seluruh *item* soal terdiri 15 butir soal tergolong valid, dengan hasil analisis realibilitas sebesar 0.717, rerata uji daya beda 0.45, semua *item* soal berada pada rentangan 0.35–0.67 pada tingkat kesukaran, validitas konstruknya *r* = 0.531, N = 100, *p* = 0.000. Uji validitas konten persentase *false negative* sebesar 6.47% dan *false positive* sebesar 9.07%. Berdasarkan hasil analisis maka instrumen layak untuk mengidentifikasi miskonsepsi siswa.

Kata kunci: sistem reproduksi manusia, miskonsepsi, *three-tier test,* siswa

INTRODUCTION

Biology is one of the sciences that is closely related to natural phenomena and has concepts that are abstract, complex and contain numerous difficult terms. It commonly leads to a difficulty among students in understanding the concept correctly. Difficulties in understanding the concept leads to misconceptions which is not in accordance with the understanding or conception agreed upon by scientists in their field (Tekkaya, 2002; Stylos et al., 2008; Suparno, 2013). In Biology, misconceptions occur, for example in genetic concept (Tsui & Treagust, 2010), on plant growth (Lin, 2004); on cell division and reproduction topic (Sesli & Kara, 2012); on the concept of meiosis (Kalas et al., 2013); and in cell biology concept(Shi et al., 2010).

Misconception analysis needs to be conducted to obtain information or input in an effort to reduce or prevent the emergence of misconceptions in the future learning. Understanding of concepts that are incorrect or misconceptions must be avoided, since the concepts contained in a material are related to future learning discussion (Kizilcik et al., 2015). Hidayati, et al. (2013) also stated that misunderstanding of concepts (misconceptions) that occur in students can interfere with students' understanding in the subsequent learning process, thus teachers need to identify students' misconception in understanding the concept. Therefore, appropriate follow-up can be given by conducting a diagnostic test identification of misconceptions. Diagnostic tests can be used to find out student weaknesses thus they can be used as a basis for follow-up (Ministry of National Education, 2007).

To identify misconceptions in students, a three-tier diagnostic test developed by Eryilmaz & Surmeli (2002) is available as one of the approaches. This instrument proved to be very accurate in measuring stu-

dents 'misconceptions, was able to observe the effectiveness of learning, was able to measure students' understanding, measure the percentage of students who did not understand the concept, and was able to distinguish misconceptions accompanied by the two previous questions (Pesman & Eryilmaz, 2010). The three-tier diagnostic instrument used consists of three layers, the first tier in the form of multiple choice questions along with the answer option, the second tier is the reason for the first tier, and the third tier is the level of confidence. Eryilmaz & Surmeli (2002) stated that the three-tier test is a combination between two tier and certainty of response index (CRI). The three-tier test is able to identify the percentage of false positive and false negative misconceptions as well as distinguising between misconception and not understanding the concept (lack of knowledge).

Misconception within student can lead to several prominent issues, particularly in human reproduction system issue. In the near future, if students have insufficient understanding it is feared that students will ignore sexual health issues (sexual transmission disease in particular) and unwanted pregnancy. Indeed, such issues are the most apprehensive issue in this current period. This is reinforced by data from the Women's Empowerment and Family Planning Agency (2009) that the number of adolescents aged 10–19 years in Indonesia is 43 million or 19.61% of Indonesia's total population of 220 million. About 1 million young men (5%) and 200 thousand young women (1%) openly state that they have had sexual intercourse. Based on the aforementioned issuses, this development research was carried out to develop and validate a three-tier test instrument to identify misconceptions of XI Graders of Senior High School on human reproductive system concept.

METHOD

The development of the three-tier test instrument was developed by adapting the Treagust (1988) model which has three main parts, the first step of defining content, followed by the next step which is obtaining information, and the last step is the development of a diagnostic test. The following is a description of the three main stages in the research development of a three-tier test instrument on reproductive system material.

Defining Content

At this stage, it began with mapping based on Campbell's Biology Book, (2016) and the 2013 curriculum syllabus on human reproductive system topic. After developing the mapping, it determined the concept proposition to be used as an instrument items with the indicator as presented in the following Table 1.

Obtaining Information

Interview to Students

Students were interviewed with several questions to find out answers and explanations from students. The results of the interview can be chosen as an answer to the first level question

Developing Open-Ended Multiple Choice Items

Open-ended or multiple-choice open test instruments was compiled based on the results of student interviews in the form of tests. Multiple choice instruments open reasons is in the form of a statement and

Table 1. Human Reproduction System Indicator Items

Question Number	Indicator
1	Analyzing the importance of circumcision for men
2	Analyzing the function of male reproductive organs
3	Analyzing the role of mitochondria in spermatozoa
4	Explaining the function of female reproductive organs
5	Analyzing the process of gametogenesis
6	Analyzing the content in urine for early pregnancy tests
7	Analyzing the causes of non-menstruation during pregnancy
8	Analyzing the occurrence of menstruation in women
9	Analyzing the occurrence of menopause in women
10	Analyzing the occurrence of giving birth
11	Describing Hormonal contraceptive methods (birth control pills)
12	Describing IUD (Intrauterine Device) / Spiral contraceptive methods
13	Analyzing the useful ingredients in ASI as antibody substances
14	Analyzing the importance of exclusive breastfeeding to infants
15	Explaining the process of "IVF" (in vitro fertilization)

have five choices of answers and students wrote reasons for answers that are freely chosen according to the students' understanding. The trial of this instrument aimed to obtain students' answers that will be used as an alternative reason for the second tier of the third-tier diagnostic instruments to be developed.

Developing a Diagnostic Test

Developing-Three-Tier Diagnostic Test

Analyzing student answers in working on multiple choice instruments for open-ended reasons will be used as an alternative reason for the second tier in the three-tier diagnostic instrument with reasonable and appropriate answer criteria.

Validating Three-Tier Diagnostic Test

Validation of questions was carried out by three validators who are experts on Reproductive Systems. The results of expert validation were used to revise the questions before being tested.

Revising Three-Tier Diagnostic Test

At this stage, a revision or refinement of the test questions was made on errors in the questions based on the validator's validation results. The revised question would be trialled to the trial subjects, a number of 100 students of XI Science Class at MAN 4 Banjar, MA Izharil Ulum, and MA Pangeran Antasari, Banjar Regency, South Kalimantan Province with purposive sampling techniques that have studied reproductive system topic. After being tested and obtained the results of reliability, validity, and the level of difficulty of the instrument and the power of difference, the questions that are not feasible will be revised or discarded and the appropriate ones will be implemented in the school.

The collected data was then analyzed descriptively by examining the answers of students referring to the answer key, then analyzing the answers between multiple choice results, reasons and beliefs of students according to the category of level of understanding in the three-tier test which is to understand concepts, guessing (lucky guessing), misconceptions, lack of knowledge, false negatives, and false positives as presented on Table 2.

RESULTS

Validity testing using the pearson correlation showed that 100% (15 items) obtained a p-value of < 0.05, thus all item items are classified as valid. A summary can be seen in Table 3.

The results of descriptive statistical analysis of score-3 (All Tier/AT) indicate that the maximum score was 13 and the minimum score was 0. The mean value of 8.26 out of 15 indicates that the average is still classified as moderate. The results of the reliability analysis with Cronbach alpha coefficient account for 0.717.

Level of difficulty analysis was testing the proportion of tier that answered correctly. Level of difficulty is displayed with index 0-1, index value > 0.90 which indicates very easy, and index value < 0.30 which indicates difficult. The results of the analysis (as seen in Table 4) show that all the question items are in the range of 0.35–0.67, meaning that they are classified as moderate. The average index value of 0.55 shows that this test is classified as having a moderate level of difficulty for students.

Discrimination item analysis between the correct response and the total score was using *point biserial correlation coefficient*. It is presented between -1 and +1, and only questions with an index > 0.20 are accepted. The results of the analysis show that 100% (15 items) have a discriminant index > 0.20. The average biserial correlation coefficient test (0.45) indicates that the test can distinguish effectively between students who have high and low conceptual understanding. The results of the analysis of the different power tests in brief are presented in Table 5.

Table 2. Three-Tier Test Conceptual Understanding Criteria

No.	*Tier* 1	*Tier* 2	*Tier* 3	Category
1	Correct	Correct	Sure	Understanding
2	Correct	Correct	Unsure	Guessing (lucky guess)
3	Correct	Inorrect	Sure	false positive
4	Correct	Inorrect	Unsure	Do not understand (lack of knowledge)
5	Inorrect	Correct	Sure	false negative
6	Inorrect	Correct	Unsure	Do not understand (lack of knowledge)
7	Inorrect	Inorrect	Sure	Misconeption
8	Inorrect	Inorrect	Unsure	Do not understand (lack of knowledge)

(Sumber: Pesman & Eryilmaz, 2010:209)

Table 3. Summary of Validity of Reproductive System Questions with Pearson Correlation

Items	Significance	Remark
1	0.003	Valid
2	0.002	Valid
3	0.001	Valid
4	0.000	Valid
5	0.020	Valid
6	0.000	Valid
7	0.000	Valid
8	0.000	Valid
9	0.000	Valid
10	0.000	Valid
11	0.000	Valid
12	0.001	Valid
13	0.000	Valid
14	0.000	Valid
15	0.004	Valid

Table 4. Summary of Index of Difficulties in Reproductive System Questions

Items	Index of Difficulties	Remark
1	0.61	Moderate
2	0.35	Moderate
3	0.42	Moderate
4	0.67	Moderate
5	0.50	Moderate
6	0.44	Moderate
7	0.65	Moderate
8	0.62	Moderate
9	0.64	Moderate
10	0.48	Moderate
11	0.50	Moderate
12	0.51	Moderate
13	0.72	Moderate
14	0.57	Moderate
15	0.58	Moderate
Avg	0.55	Moderate

Construct validity can be seen from the correlation between score-2 (Both Tier) and confidence level (Three Tier). The results show that both have strong and significant positive correlations ($r = 0.531$, $N = 100$, $p = 0.000$). The correlation between BT and TT is illustrated in the scattergram in Figure 1.

Content validity also needs to be re-confirmed with the calculation of the percentage of false negative and false positive. It is recommended that the percentage of false negatives and false positives do not exceed 10%. In this study, it was found that the percentage of false negatives was 6.47% and false positives was 9.07%. The summaries of false positives and false negatives are presented in Table 6. Based on these results, false negative and false positive values are still in the range of recommended values indicating that this test valid and can be used to identify student misconceptions.

DISCUSSION

Misconceptions need to be avoided hence they do not have a significant impact on students to further learning. Therefore, it is necessary to firstly find or prepare a good instrument to diagnose or identify student misconceptions. Therefore, the development of the three-tier test instrument on human reproductive system was developed to determine the level of student misconceptions. Based on the analysis that has been carried out, all items in question consisting of 15 items are classified as valid. In addition, the question is considered to be good if the p-value is smaller than α (at the significance level $\alpha = 0.05$), $p < 0.05$. The Cronbach alpha coefficient is 0.717, which means that the data have shown that the reliability of the reproductive system test is high. Wuttiprom et al. (2009) state that the minimum Cronbach's Alpha level of reliability is 0.70. All question items are in the range of

Table 5. Summary of Index of Discrimination in Reproductive System Questions

Items	Index of Discrimination	Remarks
1	0.294	Accepted
2	0.306	Accepted
3	0.321	Accepted
4	0.742	Accepted
5	0.232	Accepted
6	0.568	Accepted
7	0.550	Accepted
8	0.489	Accepted
9	0.605	Accepted
10	0.497	Accepted
11	0.445	Accepted
12	0.328	Accepted
13	0.463	Accepted
14	0.635	Accepted
15	0.283	Accepted
Avg	0.450	Accepted

Table 6. False Negative and False Positive in Reproductive System Questions

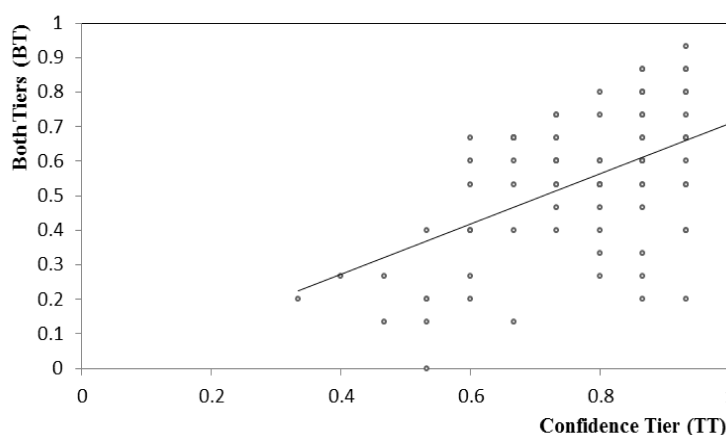| Items	False Positive (%)	False Negative (%)
1	9	4
2	8	7
3	10	4
4	11	5
5	9	7
6	10	7
7	9	7
8	9	5
9	10	5
10	6	9
11	7	10
12	12	10
13	9	3
14	11	4
15	6	10
Avg	9.07	6.47

Figure 1. *Scattergram of Score-***2** (BT) *versus confidence levels* (TT) **on the Items**

0.35–0.67 in the difficulty level test, meaning that the problem level is relatively moderate for students. The questions used must have difficulty levels ranging from 0.31–0.90 (Wuttiprom et al., 2009).

The tests have been able to distinguish effectively between students who have high and low conceptual understanding based on calculations of discrimination tests. Milencovic et al. (2016) and Taslidre

(2016), the question criteria of having a discrimination index is good if the value of D is ≥ 0.2. This instrument was also tested for construct validity which had the result that both there were strong and significant positive correlations (r = 0.531, N = 100, p = 0.000). Milencovic et al. (2016) and Taslidre (2016), Three tiers are said to be valid if there is a strong correlation between 2-Score and level of trust (p < 0.05, r > 0.5). Content validity testing needs to be done thus the instruments developed are more valid. The results of the data show that false negative and false positive values are still in the recommended range of not more than 10% which indicates that this test is valid and can be used to identify student misconceptions (Taslidre, 2016). Yet, the instruments that have been developed have several limitations, namely the questions developed are only limited to the material of the human reproductive system and the identification of misconceptions is only carried out for XI Graders of Senior High School.

Based on a theoretical study, the three-tier test instrument produced has met the suitability of product capabilities with the theoretical foundation described. As this diagnostic test, it has the ability, in (1) diagnosing weaknesses in student conceptual understanding based on analysis of student answers; (2) providing feedback quickly and individually according to the mastery of the concepts of each item; (3) assisting students to improve understanding of certain concepts (Pesman & Eryilmaz, 2010). The existence of misconceptions experienced by students can influence the delivery of concepts to students who later become biology teachers. Tekkaya (2002) suggests that teachers in schools are one of the causes of misconceptions that greatly influence students' subsequent conceptions. Repi (2004) and Ergul, et al. (2011) state that the problem that often arises in students when studying biology is the difficulty of understanding biological concepts that are abstract and difficult to understand. In this case, the teacher has a very important role to reduce or even prevent misconceptions. Therefore, it is necessary to first identify misconceptions with a diagnostic test, one of which uses a three-tier instrument.

CONCLUSION

The three-tier test instrument that has been developed to identify students' misconceptions in human reproductive system is appropriate and valid to be used to identify or diagnose misconceptions of XI Graders of Senior High School on human reproductive system. Therefore, it can be applied and implemented to identify misconceptions of XI Graders of Senior High School who have studied the material of the human reproductive system and this instrument can be used as an alternative for teachers to know students' misconceptions hence they can reduce or even prevent students' misconception earlier so that the students' misconceptions can be reduced.

Based on the results of the study, identification still requires to be applied to students towards biology learning, namely in the material of the human reproductive system, thus the results can be used as remediation in overcoming student misconceptions in human reproductive system material and teachers can improve the teaching system to be more meaningful by explaining the concept correctly.

REFERENCES

Badan Pemberdayaan Perempuan dan Keluarga Berencana (BPPKB). (2009). *Panduan pengelolaan pusat informasi dan konseling kesehatan reproduksi remaja.* Jakarta: BPPKB

Campbell, N. A., Urry, L. A., Cain M.L., Wasserman, S. A., Minorsky, P. V., & Recce, J. B. (2016). *Biology, 11th ed.* New York: Pearson Education.

Depdiknas. (2007). *Tes Diagnostik.* Jakarta: Dirjen Manajemen Pendidikan Dasar dan Menengah.

Ergul, R., Simsekli, Y., Caliz, S., Ozdilek, Z., Gocmencelebi, S. & Sanli, M. (2011). The Effects of Inquiry-Based Science Teaching on Elementary School Students' Science Process Skills and Science Attitudes. *Bulgarian Journal of Science and Eduction Policy (BJSEP), 5,* 48–68.

Eryilmaz, A., & Surmeli. (2002). Effect of conceptual assignment and change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching. 19*(10), 1001–1015.

Hidayati, T., Nugroho, S. E., & Sudarmin. (2013). Pengembangan tes diagnostik untuk mengidentifikasi keterampilan proses sains dengan tema energi pada pembelajaran IPA terpadu. *Unnes Science Education Journal. 2*(2), 311–319

Kalas, P., O'Neil, A., Pollock, C. & Birol, G. (2013). Development of a meiosis concept inventory. *CBE Life Sci Educ.,12*(4), 655–664.

Kizilcik, H. S., Çelikkanli, N. O., & Gunes, B. (2015). Change of Physics Teacher Candidates' Misconceptions on Regular Circular Motion by Time. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education. 9*(1), 205–223

Kurniawan, Y., dan A. Suhandi. (2015). The three tier-test for identification the quantity of students' misconception on newton's first laws. *Global Illuminators. 2,* 313–319

Lin, S. W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flower plant growth and development. *International Journal of Science and Mathematics Education, 2,* 175–199.

Milenkovic, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Development of a three-tier test as a valid diagnostic tool for identification of misconceptions related to carbohydrates. *Journal of Chemical Education, 93*(9), 1514–1520.

Pesman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research, 103,* 208–222.

Repi, R.A. 2004. Profil kemampuan awal dan miskonsepsi ipa biologi siswa sma se-kota Manado. *Jurnal Pendidikan Matematika dan Sains, 1*(1), 41–47.

Sesli, E. & Kara, Y. (2012). Development and application of a two-tier multiple choice diagnostic test for high school students' understanding of cell division and reproduction. *Journal of Biological Education, 46*(4), 214–225.

Shi, J., Wood, W. B., Martin, J. M., Guild, N. A., Vicens, Q., & Knight, J. K. (2010). Diagnostic assessment for introductory molecular and cell biology. *CBE-Life Sciences Education, 9,* 453–461.

Stylos, G., Evangelakis, G.A., & Kotsis, K.T. (2008). A case study in a physics department in greece. *Themes In Science and Technology Education, 1*(2),157–177.

Suparno, P. (2013). *Miskonsepsi & perubahan konsep pendidikan Fisika.* Jakarta: Grasindo.

Taslidere, E. (2016). Development and use of a three-tier diagnostic test to assess high school students' misconceptions about the photoelectric effect. *Research in Science & Technological Education, 34*(2), 164–186.

Tekkaya, C. (2002). Misconception as barrier to understanding biology. *Hacettepe Universitesi Egitim Fakultesi Dergisi, 15,* 84–93.

Treagust, D. (1988). Development and use of diagnostic tests to evaluate student's misconceptions in science. *International Journal of Science Education, 10*(2), 159–169.

Tsui, C. Y. & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education, 32*(8), 1073–1098.

Wuttiprom, S., Sharma, M. D., & Johnston, I. D. (2009). Development and use of a conceptual survei in introductory quantum physics. *Internasional Journal of Science Education, 31*(5), 631–654.