TÜRK FEN EĞİTİMİ DERGİSİ
Yıl 17, Sayı 4, Aralık 2020

Journal of
TURKISH SCIENCE EDUCATION
Volume 17, Issue 4, December 2020

http://www.tused.org

Measuring Creative Thinking Skills of Senior High School Male and Female Students in Physics (CTSP) Using the IRT-based PhysTCreTS

Edi ISTIYONO [1], WIDIHASTUTI [2], SUPAHAR [3], Syukrul HAMDI [4]

[1] Prof. Dr., Gaduate School, Universitas Negeri Yogyakarta, Yogyakarta 55281, INDONESIA, ,ORCID ID: 0000-0001-6034-142X
[2] Dr., Gaduate School, Universitas Negeri Yogyakarta, Yogyakarta 55281, INDONESIA, ORCID ID: 0000-0001-8242-658X
[3] Dr., Gaduate School, Universitas Negeri Yogyakarta, Yogyakarta 55281, INDONESIA, ORCID ID: 0000-0002-2486-5549
[4] Dr., Gaduate School, Universitas Negeri Yogyakarta, Yogyakarta 55281, INDONESIA, ORCID ID: 0000-0003-4809-1450

Received: 13.05.2019 **Revised:** 30.12.2019 **Accepted:** 21.07.2020

The original language of article is English (v.17, n.4, December 2020, pp.578-590, doi: 10.36681/tused.2020.46)

Reference: Istiyono, E., Widihastuti, Supahar, Hamdi, S. (2020). Measuring Creative Thinking Skills of Senior High School Male and Female Students in Physics (CTSP) Using the IRT-based PhysTCreTS. *Journal of Turkish Science Education,* 17 (4), 578-590.

ABSTRACT

The study is aimed at (1) describing students' CTSP, (2) comparing CTSP between male and female students, (3) identifying the highest and lowest sub-aspects of students' CTSP. The research subjects were 534 students of the State Senior High School in Bantul Regency, obtained by a stratified random sampling technique. The testing instrument was the IRT-based PhysTCreTS and data were analysed using the Partial Credit Model (PCM) IRT technique. Findings show that: (1) The highest percentage of students' CTSP lies in the "medium" category (48%), while the "high" and "very high" categories amount to 23%; (2) The CTSP of male students have a higher percentage than those of the female students but, in term of the means, male students' abilities are lower; and (3) There are differences in the CTSP sub-aspects between male students and female students. For the high category, male students are dominant on the sub-aspects planning, expressing, and formulating while female students are dominant in planning, formulating, and expressing. For the low category, male students are dominant in finding alternatives, criticizing, and testing while female students are dominant in finding alternatives, developing, and testing. The result can be used as a ground base to enhance students' creative thinking by giving special treatment to sub-aspects which fall into the low category of the creative thinking abilities of male and female students.

Keywords: Creative thinking skill, gender difference, Senior High School physics learning.

Correspondence author e-mail: edi_istiyono@uny.ac.id

INTRODUCTION

In the 21st century, the world of science and technology has experienced fast development in various aspects of life. In relation to this fast development in the globalization era, directly or indirectly, quality of human resources is inevitably need to face with challenges. The quality of human resources can be obtained through quality education. Thus, elevating the quality of education is an important factor in determining the success and progress of nation building. The role of technology has been studied, including in the field of education. According to research by Gulen & Yaman (2019), the results suggested that integration of STEM disciplines into Toulmin's argumentation model can be used for increasing academic achievement of students, developing of the reflective thinking, and observing the development of psychomotor skills at the formation of arguments in the classrooms. A research from Çelik, Kırındı, & Kotaman (2020) showed that the computer-based analogy method was more effective than the present program applications in terms of students' academic success in science lessons and permanence of knowledge.

The surveys conducted by *Science Literacy Version Programme for International Students Assessment* (PISA) showed that Indonesia occupied position 64 in 2012 and moved to 62 in 2015 which means encountering a serious problem for Indonesian education (OECD, 2018). Besides, a 2015 survey by *Trends in International Mathematics and Science Study* (TIMSS) showed that, in science competency average, Indonesia made a score of 397 with respect to 500 of the world's average. This score placed Indonesia at the 45th rank out of 48 countries. This average score was a decrease of 9 points compared to the 2011 score which was 406 with respect to 500 of the world's average (TIMSS & PIRLS, 2015). As seen from these survey results, Indonesia has experienced a decrease in the quality of students' competitiveness in the international arena. Improvement of the quality of education has become one of the priorities of the Government by conducting a reform in the curriculum, instructional processes, and assessment that are all directed to prepare students to face challenges in the 21st century. In these efforts, improvement was made in the education system by adapting the educational objectives so that students have life skills and abilities.

Physics is a part of science that becomes a basis for advancement in science and technology. Chiappetta & Koballa (2010) stated that physics is "a way of thinking, a way of investigating, a body of knowledge, and science and its interactions with technology and society". Therefore, physics becomes an important need of learning in the school.

In physics instruction, Etkina, *et al*. (2006) stated that the students' scientific knowledge should not only involve scientific processes, but it should also develop students' skills in reflective and creative thinking. Some scientific competencies that are developed in physics education can be listed as: (1) representing a physics process in number of ways, (2) planning and testing a description qualitatively and quantitatively, (3) change a qualitative description into quantitative, (4) planning an experimental research, (5) collecting and analyzing data, (6) doing an evaluation, and (7) communicating the obtained results.

Binkley, Erstad, Herman, Raizen, Ripley & Rumble (2010) stated that the "ways of thinking" are divided into three skills: (1) being innovative and creative, (2) thinking critically in making a conclusion, and (3) learning to manage meta cognitive skills. Meanwhile, Siswono (2004) stated that creative thinking is a process that is used when deriving or exposing a new idea. This involves coordinating the ideas that has not been done before. Creative thinking is the ability to think sharply, using intuition, and moving imagination to uncover new possibilities or new ideas as a development of old ideas to solve a problem from various point of views. Creative thinking is an influential ability which students have to possess and avoid having creative anxiety. According to study conducted by Daker, Cortes, Lyons, & Green (2020), the result indicated that creativity anxiety might have wide-reaching impacts and distinguished creativity anxiety from anxiety about non-creative aspect of

performance. To obtain deeper information, it was also needed to consider in terms of gender in to order to be able to offer the best solution.

Gender relation is a social relation between men and women that is helpful for each other and that has differences and inequalities. Gender relation is different from time to time and from society to society due to differences in tribes, religions, social status, traditions, and norms (Amir, 2013). In another manner, Santrock (2002) described gender as an identity based on biological differences from birth and referring to the biological dimensions of men and women. It is possible that there are differences in the levels of creative thinking based on gender differentiation.

Cramond, Morgan, Bandalos, & Zuo (2005) showed the presence of differences in creativity levels between different genders quantitatively and qualitatively. In his study of 82 children, Aziz (2006) showed that there is a tendency that girls have a higher level of creativity than boys. From these results of studies, it can be said that there is an influence of gender differences in the levels of creative thinking. The term is needed to be analyzed more deeply in order to provide problem solving and appropriate learning methods for the gender differences. This is a strong base in conducting the research. So, the categorized aspects or sub-aspects as low or high are related to critical thinking abilities of high school students. Aiami & Hagani (2012) confirmed the importance of using new teaching methodologies to raise students' creative thinking skills. The implemented method was able to strengthen and enhance the creative thinking of students who are still in the low category.

One of the ways to reveal the progress of students' learning is by doing an assessment. An education assessment is the process of collecting and analyzing information to measure students' progresses in learning (Decree of the Ministry of Education and Culture, 2013). Furthermore, according to (Mardapi, 2012), an assessment is the collecting of data personally to obtain the characteristics of the person. Assessments can also be designed in everyday contexts and measure overall components called authentic assessments (Sabtiawan, Yuanita, & Rahayu, 2019). Authentic assessment is an alternative assessment forcing students to perform like a professional in a real work-place. In other words, this type of assessment trains students to be successful-performers in professional jobs. In physics education, assessment is an activity of collecting data individually that will give a description of each person. In physics learning, learning assessment shows the characteristics of each student's results in learning in the form of scores. Assessment is carried out using an assessment instrument in the form of a test. The test can be oral or written consisting of test items and response format of test takers.

Thus far, many assessment instruments are developed in the levels of remember in knowledge dimension including items designed to measure levels of creative thinking (Istiyono, Dwandaru & Farida, 2018). In education assessment, two measurement approaches are known: classical and modern. Suryabrata (2012) stated that the classical theory of assessment is still much used up to the present time. In the same way, the measurement of creative thinking skills is still conducted by using the classical mode. However, Widhiarso (2010) stated that the use of modern assessment techniques have the tendency to make the test takers with higher thinking skills to get higher score than the test takers with lower thinking skills. Therefore, it is often recommended that the modern test theory be used to measure creative thinking skills with a better result.

In the classical theory of measurement, test scoring is usually based on the steps when answering a test item correctly. Scoring is conducted for every step and each item score is summed to produce a raw score. This classical mode of scoring is not fully acceptable since the level of difficulty of each step is not calculated. Therefore, this classical measurement theory is perfected by the modern theory. The modern theory in measurement is also called the item-response theory. This item-response approach is an alternative technique to be used

in analyzing a test. Two principles are used in this approach: the principle of relativity and the principle of probability (Retnawati, 2014).

One of the assessment models to analyze responses for creative thinking measurement is the Partial Credit Model (PCM) (Van der Linden & Hambleton, 1997:101). The PCM is a development of the Rasch model for the dichotomous items applied for the polytomous items. The dichotomous Rasch model which has one item location parameter is developed by spreading item locations into categories. An assumption in the PCM model is that each test item has the same discriminating power but the difficulty indexes in each phase does not have to be in order. In other words, one phase may be more difficult than the next (Retnawati, 2014). The categorical score in the PCM indicates the number of phases to answer the item correctly. The PCM is used to analyze test items that require several phases to answer correctly.

Among others, multiple-choice test format is the most used in assessment activities for certain advantages such as: (1) test material is representative of the over-all instructional material, (2) scoring can be done rapidly and easily, (3) right or wrong answer helps the test to be objective (Sudjana, 1990).On the contrary, the multiple-choice test format makes it possible to do guessing so that test taker's way of thinking cannot be seen clearly (Sudjana, 1990). Therefore, another test format is needed to minimize these weaknesses. One of the alternatives is the multiple-choice test with reasons. In this test type, the test taker place their choice of the correct answer and give reasons for making the choice. Reasons given by the test taker can be used to measure their skills. Black and Willian (2010) gave a definition of an evaluation or assessment activity as one conducted by the teacher on the students to test the students' instructional performance and to give information as a feedback for improving the instructional processes. Hence, based on the modern theory of testing, a test instrument has been developed to measure creative thinking skills called the *PhysTCreTS* (Istiyono, Dwandaru & Farida, 2018).

In order to improve the students' competencies in the 21st century education, there is a need to develop students' creative thinking skills in physics (CTSP). For this, a preliminary profile of CTSP needs to be known for Senior High School (SHS) students in order to be able to develop them. Therefore, an effort is needed to measure and analyze the CTSP of SHS students, both male and female. PhysTCreTS is an instrument to measure students' creative thinking abilities in physics which had been developed by the authors and had been through a process of calibration and validation. So, it is appropriate to be used in this study.

METHODS

a) Measurement of Creative Thinking Skills

Steps of the measurement includes the following steps: 1) Packaging of the test, 2) Determining the test takers, 3) Testing, 4) Analysis of the results, and 5) Interpretation. These steps are visually presented in Figure 1.

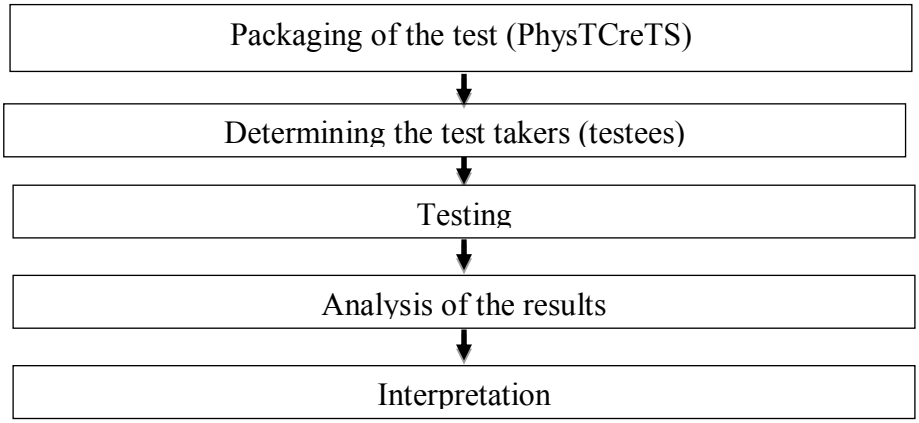

Figure 1. Steps in the creative thinking skill testing

b) Research Instrument

The used instrument was a PhysTCreTS which was successfully developed to measure creative thinking skills. The measured aspects were Fluency, Flexibility, Originality and Elaboration. To be clearer in terms of measured aspects and sub-aspects, they are summarized in the following Table 1.

Table 1. *Aspects and Sub Aspects of Measured Instrument Using PhysTCreTS*

Aspects	Sub Aspects
Fluency	Formulating the Answers
	Revealing the Idea
	Criticizing an Object
Flexibility	Doing Interpretation
	Looking for Alternative Answers
	Categorizing
Originality	Planning the Novelty
Elaboration	Solving the Problems with the Specific Procedures
	Developing the Idea
	Examining

Based on the aspects and sub-aspects above, it was then compiled into a test to measure students' creative thinking skills in physics. The test of Creative Thinking Skills was packaged into two groups, namely package A and package B. Each package was included 7 questions out of 35 to be used.

c) Validity and Reliability of the Instrument

The test of Creative Thinking Skills (PhysTCreTS) was firstly tested in quality before implementation by using the estimation test of validity and reliability. To prove the validity, content validity with the Aiken formula and reliability estimation of Cronbach alpha were utilized. In addition, the validity and reliability of the test instrument were tested to determine the quality of the considered questions based on the level of difficulty and the index of discrimination using Item Response Theory with the Partial Credit Model (PCM).

PhysTCreTS has a measure of validity and reliability and has been found feasible to measure creative thinking skills of SHS students (Istiyono, Dwandaru & Farida, 2018). Therefore, the content validity was obtained via expert judgment which analyzed through the

Aiken technique. The obtained Aiken's index V ranged from 0.71 to 0.76. And from the empirical try-out, it was found to have a fit with the PCM by MNSQ INFIT values ranging from 0.93 to 1.10. The item difficulty levels ranged from -1.06 to 1.04. All the test items had a reliability measure with a reliability score of 0.72 and the test had a fit with the students' abilities of -1.8 to 3.0 (-1,8≤θ≤3) based on the information functions and *SEM*.

d) Respondent Selection and Test Administration

Using a stratified random sampling technique, a total of 534 SHS students were taken as the research subjects. These were students of State SHS 3 in Bantul Regency with the ability levels of low, medium, and high based on the results of the physics national examination. The sample size was of the "very good" category since the IRT analysis model of PCM 1PL requires a minimum number of 250 subjects. During the test administration, seating was arranged according to the required procedure in that: (a) adjacent students took different test codes; Test A or Test B and (b) the time allocation for doing the test was 90 minutes.

e) Data Analyses

The data analyses was included the followings: 1) determining the students' creative thinking skills in physics, i.e. CTSP, 2) calculating the percentages of each level, and 3) comparing the CTSP between male and female students and determining the sub-aspects of CTSP that are dominantly high or dominantly low for male and female students.

First, in order to find the CTSP from students' responses, two analysis techniques are available; namely the classical technique and the item response theory (IRT). The classical theory model is the more popular among users. In the IRT model, the test analysis is conducted on the basis of the numbers of the test items and test takers. The IRT model used was of the 1 Parameter Logistic (1 PL). Scoring of this multiple-choice test with reasons was polytomous following the partial credit model (PCM).

The PCM is formerly a dichotomous Rasch Model that is extended into polytomous items. The one-parameter logistic model (1 PL) is extended through breaking the item location into a number of categories. If *i* is a polytomous item with score categories 1, 2, 3, …, *m*, the probability of the individual *n* score *x* on item *i* figured out in the category response function (CRF) is formulated as in Equation 1 (Ostini & Nering, 2006: 28; Muraki & Bock, 1997).

$$P_{ig}(\theta) = \frac{\exp\left[\sum_{g=o}^{l}(\theta - b_{ig})\right]}{\sum_{h=o}^{m}\exp\left[\sum_{g=o}^{h}(\theta - b_{ig})\right]} \qquad (1)$$

Notes:

$P_{ig}(\theta)$= the probability the test taker with ability level θ obtains score category on item -i^{th}

θ = individual trait level (individual trait location on continuum of latent trait or ability)

b_{ig} = item location parameter or difficulty level (showing the probability of obtaining score 0 or 1 is equal)

Second, the percentage of each level is done by determining the levels of "very low", "low", "medium", "high", and "very high" with the ideal means and ideal standard deviations presented in Table 2. The calculation of the ideal mean (M_i) and standard deviation (SB_i) is based on the highest and lowest scores of the research variable as shown in Equations 2 and 3 (Azwar, 1998).

Table 2. *Score Interval of Ability Level with*

No	Interval of Ability	Level
1	$M_i + 1.5SB_i < \theta$	Very High
2	$M_i + 0.5SB_i < \theta \leq M_i + 1.5SB_i$	High
3	$M_i - 0.5SD_i < \theta \leq M_i + 0.5SD_i$	Medium
4	$M_i - 1.5SD_i < \theta \leq M_i - 0.5SD_i$	Low
5	$\theta < M_i - 1.5SD_i$	Very Low

$$M_i\,(ideal\ mean) = \frac{1}{2}(\theta_{max} + \theta_{min}) \quad (2)$$

$$SD_i(ideal\ standard\ deviation) = \frac{1}{6}(\theta_{max} - \theta_{min}) \quad (3)$$

Third, the CTSP of male and female students were regrouped and compared by the highest and lowest abilities, means, and standard deviations. Subsequently, determination was done on the sub-abilities dominantly high and dominantly low for male and female students.

FINDINGS

a) Creative Thinking Skill in Physics

Results of the skill assessment were presented via a score range from very low to very high. This was done by providing score categories. The creative thinking test used in the study had a maximum $\theta$of 4 and a minimum θ of -4. Thus, with an ideal mean of 0 and five standard deviations for the five range categories, an ideal standard deviation of 1.33 was obtained.

The results of the CTSP testing for all the test takers were plotted in five levels as can be seen in Table 3.

Table 3. *Percentage Levels of CSTP*

CSTP Level	Percentage (%)
Very Low	6
Low	17
Medium	48
High	25
Very High	4

As can be seen in Table 3, the dominant Berdasarkan CSTP was in the "medium" level (48%) while the one above "high" and "very high" amounts to 29%. This also showed that students' CTSP was at the medium up to high category.

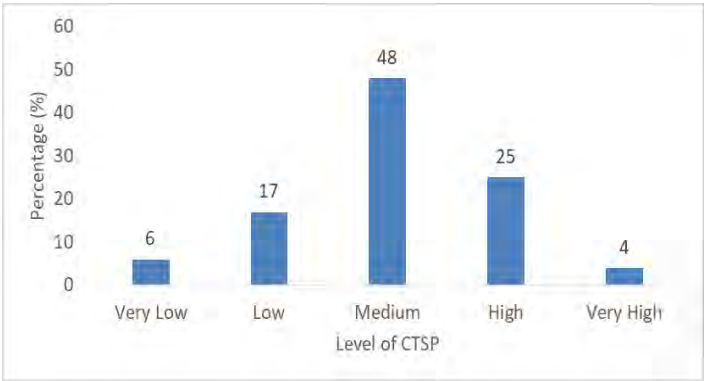

Figure 2. CSTP Percentage Levels

Figure 2 shows that the students' CTSP had a normal distribution pattern both for the lowest percentage and the highest (4%). This is due to the fact that the "very high" level needs students who really have a very good understanding of the material being tested. On the other hand, the largest proportion of the students was in the medium category (48%). Percentages of the "medium", "high", and "very high" summed up and a percentage of 77% was obtained. This means that students' abilities are medium-up (above average).

Then, in Figure 3, the dominant sub-aspects of the CSTP, lying within the high categories (3 and 4) were planning, formulating, and expressing. Meanwhile, the dominantly low CSTP sub-aspects (1 and 2) were "among others", "finding alternatives", "criticizing", and "testing".

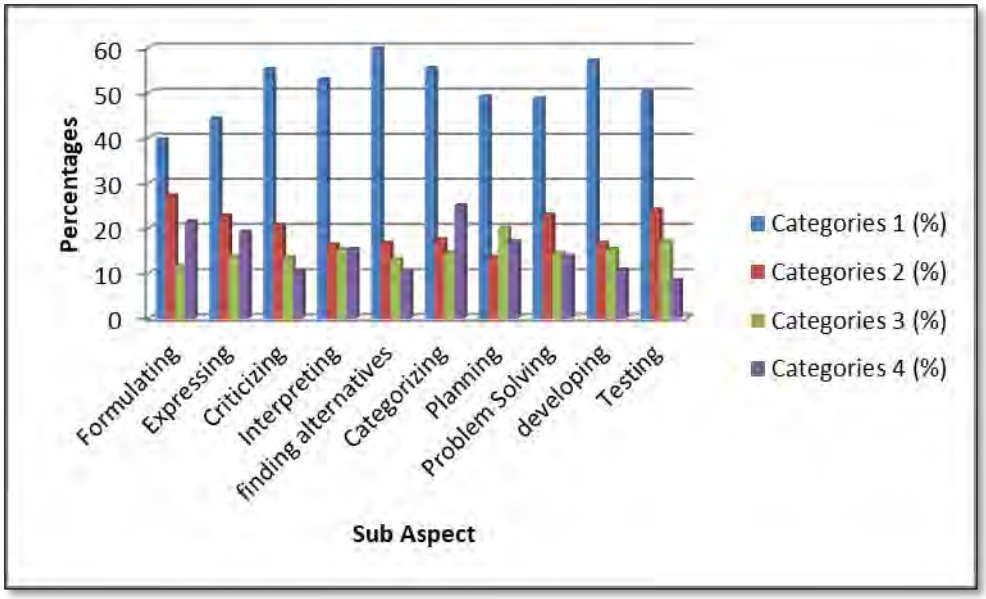

Figure 3. Students' CTSP Categories (1 to 4)

b) Creative Thinking Skill in Physics between Male and Female Students

Results of the calculation of CTSP levels in terms of male and female students can be seen in Table 4. For the male category, the percentage of the "medium" up to "very high" levels consisted 80% of the total. This means that the male students have a CTSP level above the "high" category.

Table 4. *CTSP Levels by Male and Female Students*

CTSP Category	Male (%)	Female (%)
Very Low	7	5
Low	13	20
Medium	51	44
High	27	27
Very High	2	4

Meanwhile, for the female students, the added percentage comes to 71%. This means that the female students have a CTSP level of above the average.

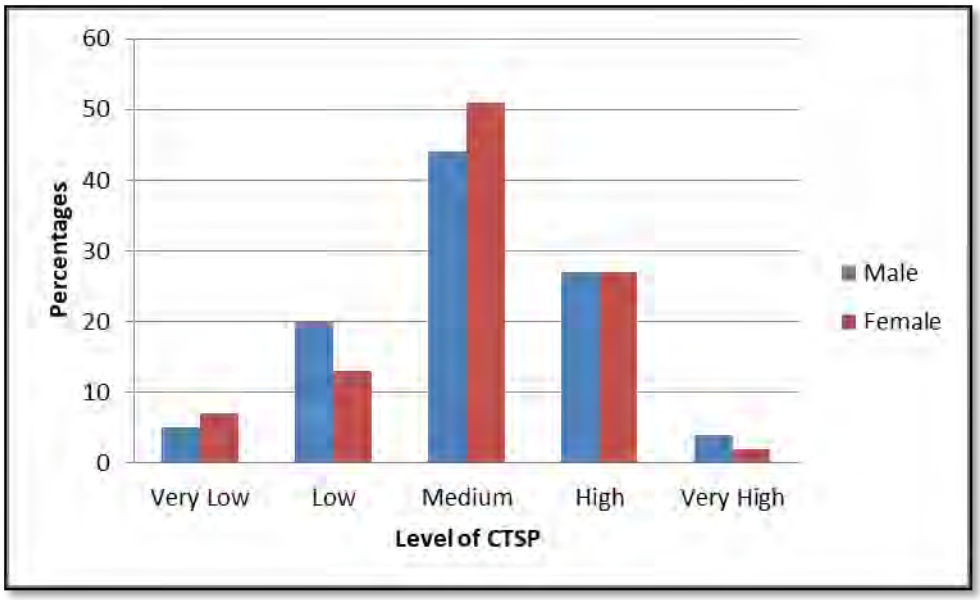

Figure 4. CTSP Percentages of Male and Female Students

From Figure 4, it can be seen that there was a slight difference of CTSP percentages between male and female students in terms of the three high levels, male = 80% and female = 75%. The male students noted a higher percentage of 5% than female. However, from Table 5, it can be seen that female students are 0.02447 higher than male students on the Logit scale. On this scale, female students were higher than male.

Table 5. *CTSP Measures of Male and Female Students by Maximum Value, Minimum Value, Mean, and Standard Deviation*

CSTP Value	Male (Logit Scale)	Female (Logit Scale)
Maximum	1.66360	2.26830
Minimum	-3.71900	-4.43720
Mean	-0.01443	0,01004
SD	0.73926	0.67840

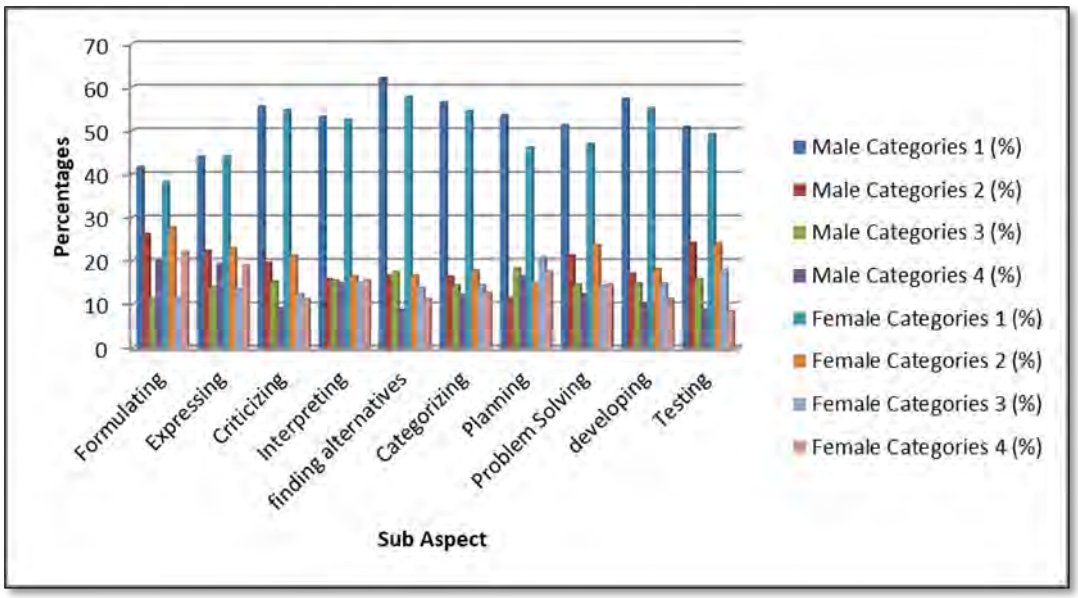

Figure 5. CTSP Percentage Categories (1 to 4) of Male and Female Students

Figure 5 shows that three dominant CTSP sub-aspects (3 and 4) of male students were planning, expressing, and formulating while those for female students were planning, formulating, and expressing. For the dominantly low scores (1 and 2), male students noted finding alternatives, criticizing, and testing while female students noted finding alternatives, developing, and testing.

DISCUSSION

Research results showed that there is a difference in creative thinking skills for physics between male and female students. This is neithher in concord with Heong, Othma, Yunos, Kiong, Hassan & Mohamad (2011) nor with Reese *et al*. (2009) who reported in their studies that differences in gender do not have a significant influence for males and females in their abilities of divergent thinking.

However, results are in agreement with a number of previous other studies. Edgerton, Tracey & Lance (2014) reported that there is a significant influence of gender differences in certain aspects. Students' learning achievements are also affected by economic conditions of the family. Slameto (1991) stated that the condition of the family's economy is tightly related to children's learning achievements. Family's economic conditions have a relatively stronger influence on daughters than on sons. Besides, children's habits also affect academic achievements. Boys' habits are believed to be effective on school achievements than girls' habits. In addition, Tanti et.al (2018), stated that teachers or lecturers must pay attention to self-regulated strategies in studying physics to improve student achievement in learning physics. This is confirmed by the results of the study Sukardiyono, Rosana, D. & Dwandaru, W. S. B. (2019) that the students' science process skills (SPS) levels were very good for predicting, experimenting, observing, and measuring, while 'concluding and communicating' aspects were good.

Zubaiah, Nur, Susriyati & Endang (2017) stated there are differences in creative thinking skills between male students and female. Furthermore, that male students are believed to have higher creative thinking skills than female students due to the differences in the brain anatomy. The brain anatomy is believed to carry differences in male and female students in their learning patterns and activities. Mahanal's study, of SHS students in Malang City in 2011, reported differences between male and female students. For example, female students are found to have a higher critical competency than males. In a stronger tone, Walsh & Hardy (1999) reported that female university students have higher competencies in all aspects of critical thinking than male students. These research findings show that female students have a higher level of thinking skills than male students. One explanation for this difference is that female students are thought to be superior in the use of language. Other studies reported that female students are more proficient than male students in their verbal competencies (Elliot *et al*., 2000; Sasser, 2010). Language can be regarded as a medium for expressing one's thoughts. De Bono (1990) also emphasized that language or verbal symbol systematic is important as a medium of thinking, perhaps even the most, such that thinking is often called as a soul dialogue. And Masoud Mahmoodi-Shahrebabaki (2015) stated that language influences creativity. In another study, Jonni Sitorus (2017) stated that students' creative thinking competencies can be stimulated by having innovations in the classroom instruction such as using Realistic Mathematics Education (RME). Meanwhile, Marc Spooner (2002) explained that creative products are derived from a large number of social influences. Furthermore, it is emphasized relations among students, school, and social communities becoming an inter-responsibility that is quite influential in the successful products resulting out from creative attempts.

CONCLUSION

As a follow-up of the research findings, items of conclusion can be presented as follows. First, the highest percentage of students' CTSP levels is at the "medium" category (48%), "high" and "very high" categories are represented by 23%. Second, CTSP abilities of male students are different from females: (a) on percentage levels, male students' abilities are higher than females and (b) based on means, male students' abilities are lower than females. Third, differences in the three dominant CTSP sub-aspects between male and female students are as follows: (a) on the high category, male students are dominant in the sub-aspects planning, expressing, and formulating while female students are dominant in planning, formulating, and expressing. and (b) on the low category, male students are dominant in finding alternatives, criticizing, and testing while female students are dominant in finding alternatives, developing, and testing. The result will be used as a ground base for the process of physics learning improvement to boost students' creative thinking by providing an appropriate treatment for male and female students.

REFERENCES

Amir, Z.M.Z. (2013) Perpektif Gender dalam Pembelajaran Matematika [Gender Perspectives in Mathematics Learning]. *Jurnal Marwah Vol. XII No.1 Juni Th. 2013*. page 15-31

Aziz, R. (2006). Studi tentang kreativitas pada siswa Sekolah Menengah Pertama di kota Malang [Study of creativity in junior high school students in the city of Malang]. *Psikoislamika*, 3, 3, 239-254.

Azwar, S. (2012). *Reliabilitas dan validitas [Reliability and validity]*. Yogyakarta: Pustaka Pelajar.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2010). *Defining 21st century skills*. Australia: University of Melbourne.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81-90.

BSNP. (2006). *Penyusunan KTSP Kabupaten/ Kota; Panduan Penyusunan Kurikulum Tingkat Satuan Pendidikan Jenjang Pendidikan Dasar dan Menengah* [Preparation of Regency/City KTSP; Guide to Compiling Curriculum Levels of Primary and Secondary Education]. Jakarta: Departemen Pendidikan Nasional.

Chiappetta, E. L and Koballa, T. R. (2010). *Science Instruction in the Middle and Secondary Schools.* New York: Allyn & Bacon.

Cramond, B., Morgan, J.M, Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrence tests of creative thinking: Alive and Well in the new millennium, *Gifted Child Quarterly*, 49, 4, 283-291.

Çelik, H, Kırındı, T. & Ayçiçek Kotaman, Y. (2020). The Effect of the Computer-Based Analogy Used in Science Teaching on Learning. *Journal of Turkish Science Education, 17*(1), 73-93.

Daker, R. J., Cortes, R. A., Lyons, I. M., & Green, A. E. (2020). Creativity anxiety: Evidence for anxiety that is specific to creative thinking, from STEM to the arts. *Journal of Experimental Psychology: General, 149*(1), 42–57. https://doi.org/10.1037/xge0000630

De Bono, E. (1990). Mengajar Berpikir [Teaching Thinking]. Translater: Soemardjo. 1992. Jakarta: Erlangga.

Edgerton, Jason D, Tracey Peter, Lance W. Robert. (2014). Gendered Habitus and Gender Differences in Academic Achievement. *Alberta Journal of Education Research*, Volume 60(1), 182-212.

Elliot, S. N., Kratochwill, T. R., Cook, J. L. & Travers, J. F. (2000). *Educational Psychology: Effective Teaching, Effective Learning*, Third Edition. United States of America: McGraw-Hill Companies, Inc.

Etkina, E., Alan van H., Suzanne W. B., David T. B., Michael G., Sahana M., David R. & Aaron W. (2006). Scientifics abilities and their assessment. *Physical Review Special Topics-Physics Education Research*2, 020103: 1-15.

Gülen, S. & Yaman, S. (2019). The Effect of Integration of STEM Disciplines into Toulmin's Argumentation Model on Students' Academic Achievement, Reflective Thinking, and Psychomotor Skills. *Journal of Turkish Science Education, 16*(2), 216-230.

Heong, Y. M., Othman, W.B., Yunos, J.B.M., Kiong, T.T., Hassan, R.B., & Mohamad. M.M.B. (2011). The Level of Marzano Higher Order Thinking Skills among Technical Education Students. *International Journal of Social Science and Humanity*, Vol. 1, No. 2, July 2011, (Online), (http://www.ijssh.org/papers/20-H009.pdf), It was acceded on 26 November 2017.

Istiyono, Edi, Wipsar Brams Dwandaru, & Farida Rahayu. (2018). Pengembangan tes creative thinking skills fisika SMA (*PhysCreTHOTS)* berdasarkan teori tes modern [Development of high school physics creative thinking skills tests (PhysCreTHOTS) based on modern test theory]. *Cakrawala Pendidikan*, Th. XXXVII, No. 2, 2018.

Mahanal, S. (2011). Pengaruh Pembelajaran Berbasis Proyek pada Mata pelajaran Biologi dan Gender terhadap Keterampilan Metakognisi dan Kemampuan Berpikir Kritis Siswa SMA di Malang [The Effect of Project Based Learning on Biology and Gender Subjects on Metacognition Skills and Critical Thinking Ability in High School Students in Malang]. *Research Report*. Malang: Lemlit UM.

Mahmoodi, Masoud & Shahrebabaki. (2015). Relationships Between Language Teachers' Time-Management Skill, Creativity, and Burnout: A Mediation Analysis. *Alberta Journal of Education Research*, Volume 61(1), 20-39.

Mardapi, Djemari. (2012). *Penyusunan intrumen tes dan nontes* [Preparation of test and non-test instruments]. Yogyakarta: Mitra Cendekia

M. L. Nering & R. Ostini. (2006) *Handbook of Politomous Item Response Theory Models*. London

Muraki, E., & Bock, R. D. (1997). *Parscale 3: IRT based test scoring and item analysis for graded items and rating scales:* Chicago: Scientific Software Inc.

Reese, H. W., Lee, L., & Cohen, S. H., Puckett, J. M. (2009). Effects of intellectual variables, age, and gender on divergent thinking in adulthood. *International Journal of Behavioral Development*, 2001 25 (6) 491–500, (Online), (nccuir.lib.nccu.edu.tw/bitstream/140.119/20791/1p99.pdf), It was acceded on 26 November 2017.

Retnawati, Heri. (2014). *Teori Respons Butir dan Penerapannya* [*Item Response Theory and Its Application*]. Yogyakarta: Nuha Medika

Santrock. W. J. (2002) *Life-Span Development 2,* Translater: Chusairi & Damanik). Jakarta: Erlangga

Sitorus, Jonni. (2017). Students' Thinking Way: Learning Trajectory of Realistic Mathematics Education. *International Journal of Innovation in Science and Mathematics*. Volume 5(1), 22-31.

Slameto. (1991). *Belajar dan faktor yang Mempengaruhinya* [Learning and factors that influence]. Jakarta: Rineka Cipta

Spooner, Marc. (2001). Creative Teenage Students: What Are They Telling Us About Their Experiences in (and Around) Our High Schools? *Alberta Journal of Educational Research*, Volume 48(4), 314-326.

Suryabrata, S. (2002). *Pengembangan Alat Ukur Psikologis* [Development of Psychological Measurement Instrument]. Yogyakarta: Andi Offset.

Widhiarso, Wahyu (2010). *Model Politomi dalam Teori Respons Butir* [Politomous Model in Item Response Theory]. Yogyakarta: Psikologi UGM.

Van der Linden, Wim J and Hambleton, Ronald K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag New York, Inc

OECD. (2015). PISA 2015 Result in Focus [Online] https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf. It was acceded on May 2018

TIMSS & PIRLS. (2015). International Result in Science [Online] http://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/full%20pdfs/T15-International-Results-in-Science-Grade-8.pdf. It was acceded on 19 May 2015

Sabtiawan, W. B., Yuanita, L., & Rahayu, Y. S. (2019). Effectiveness of Authentic Assessment: Performances, Attitudes, and Prohibitive Factors. *Journal of Turkish Science Education, 16*(2), 156-175.

Sasser, L. (2010). Brain Differences between Genders. Gender Differences in Learning, Genesis 5:12, (Online), (http://www.faccs.org/assets/ Conventions/Convention 10/Workshops/SasserGenderDifferences-in-Learning.pdf),

Sudjana, N. (1990). *Penilaian Hasil Belajar Mengajar* [Teaching and Learning Outcomes Assessment]. Bandung: PT Remaja Rosdakarya.

Sukardiyono, Rosana, D. & Dwandaru, W. S. B. (2019). Measuring junior high school students' science learning and science process skills through and integrated science instructional assessment. *Journal of Turkish Science Education, 16* (4), 467-477.

Tanti, Maison, Mukminin, Syahria, Habibi, & Syamsurizal, (2018) Exploring the Relationship between Preservice Science Teachers' Beliefs and Self-Regulated Strategies of Studying Physics: A Structural Equation Model. *Journal of Turkish Science Education, 15*(4), 79-92

Walsh, C. M., & Hardy, R.C. (1999). Dispositional Differences in Critical Thinking Related to Gender and Academic Mayor. *Journal of Nursing Education*, Apr 1999; 38, 4; ProQuest Central pg. 149, [Online]. http://textos.pucp.edu.pe/textos/descargar/1111.pdf, It was acceced on 26 November 2017.

Zubaidah, Siti, Nur Miftahul Fuad, Susriyati Mahanal, Endang Suarsini. (2017). Improving Creative Thinking Skills of Students Through Differentiated Science Inquiry Integrated with Mind Map. *Journal of Turkish Science Education*, Volume 14(4), 77-91.