Tadris: Jurnal Keguruan dan Ilmu Tarbiyah 7 (2): 241-258 (2022)
DOI: 10.24042/tadris.v7i2.13265

TADRIS: JURNAL KEGURUAN DAN ILMU TARBIYAH
(Tadris: Journal of Education and Teacher Training)
P-ISSN: 2301-7562 | E-ISSN: 2579-7964
ejournal.radenintan.ac.id/index.php/tadris/index

Instrument Analysis of Biology Teachers' Needs to Assess Students' Creative Thinking Skills Using RASCH Model

Rezqi Rohmatika Sari, Bowo Sugiharto*, Sri Widoretno
Master Program of Biology Education, Universitas Sebelas Maret, Indonesia

Article History:
Received: July 31st, 2022
Revised: August 31st, 2022
Accepted: September 21st, 2022
Published: December 29th, 2022

Keywords:
Assessment,
Creative thinking skill,
RASCH model

***Correspondence Address:**
bowo@fkip.uns.ac.id

Abstract: This study aimed to investigate the validity, reliability, scale understanding, item difficulties, and bias items of the instrument for analyzing the needs of biology teachers regarding the assessment of creative thinking skills using the RASCH model. The instrument was modified based on the indicators of divergent and convergent thinking processes in the scientific learning stages integrated with the dimensions of creative thinking skills by using ADDIE research model. The research participants comprised of 104 participants from the Solo Raya area, those of whom were 64 Senior High School Biology teachers and 40 Junior High School Science teachers. The research instrument used inventory with a Likert scale of 1 to 4. The instrument validity, reliability, scale understanding, item difficulties, and bias items were analyzed by applying the RASCH model using Winstep 3.73. The research results showed that the overall validity was acceptable, and the item validity did not need improvement. Overall reliability was very good, and the item reliability was excellent. Rating scale analysis showed respondents had understood the Linkert scale of 1 to 4 well. Based on the item difficulties results of the teachers' responses, it was found that there had not been any teachers who evaluated the indicators of creative thinking skills by using writings or pictures in problem-solving assessment. The bias test results on the instrument items indicated that five items could potentially be biased due to age difference, and the two others were due to gender type. Therefore, the development of assessment instruments for creative thinking skills with scientific and social problem-solving based assignments, as well as writing and visual expressions, is required.

INTRODUCTION

Creative thinking skills are the ability in creating, implementing, communicating, and working creatively with others (Tran et al., 2017; Trilling & Fadel, 2009). The developing of people's creative thinking skills has actually become one of the educational goals in the 21st Century and, therefore, it is also essential to evaluate creative thinking skills (Lucas, 2016; Sugiharto et al., 2019; Tran et al., 2017). One of Several international institutions that much concern to evaluate creative thinking skills is the Global Innovation Index (GII) as a global-level measuring organization for innovation (Dutta et al., 2020). Based on the results of the GII evaluation in 2022, Indonesia places the second quarter, ranked 87th out of 132 countries (Dutta et al., 2020). Such result indicates that the scores of the Indonesian institution, human resources and research, business sophistication and innovative products are

below the average (Dutta et al., 2020). These results show that the human resources' creative thinking skills to innovate in Indonesia are still low (Dutta et al., 2020).

Due to the importance of measuring creative thinking skills, the need analysis of Biology teachers to assess creative thinking skills is essential. The indicators of creative thinking skill assessment can be used as references to recognize someone's creative thinking skills (Sarriot et al., 2014; Yustina et al., 2020). The indicators of fluency, flexibility, originality, and elaboration (FFOE) have long been used to measure an individual's creative thinking skills. The FFOE indicator focuses on divergent thinking processes, which is a thinking process involving imagination in creating innovation (Guilford, 1975; Jia et al., 2017; Oppezzo & Schwartz, 2014; Runco et al., 2017; Runco & Acar, 2012). Fluency is the ability to mention as many relevant ideas as possible (Guilford, 1975; Hass, 2015; Runco & Albert, 1985; Zhou et al., 2020). Flexibility is distinguishing and classifying ideas from different points of view (Jasim Mohammed & Ati Daham, 2021; Lia D Rubenstein et al., 2019). Originality is the novelty of the idea or product that has been created (Bart et al., 2017; Guilford, 1975; Hass, 2015). Elaboration is the ability to detail the construct of ideas until a solution is found (Bart et al., 2017; Oppezzo & Schwartz, 2014; Lia D Rubenstein et al., 2019; Runco et al., 2017).

Creative thinking skills require the ability to evaluate ideas from divergent thinking processes (Barbot & Lubart, 2012; Catarino et al., 2019; Hass, 2015; Vally et al., 2019) and convergent thinking processes which is the ability to evaluate (Charyton et al., 2011; Oppezzo & Schwartz, 2014). The indicators of convergent thinking processes found in creative thinking skills involve the usefulness, evaluation, and improvement (Benedek et al., 2006; Shu-Chen et al., 2020; Vally et al., 2019). The usefulness is the ability to explain the utility of new ideas or products (Benedek et al., 2006; Charyton et al., 2011; Lia D Rubenstein et al., 2019). The evaluation is the ability to evaluate the advantages, disadvantages, and possibilities of implementing new ideas and products (OECD, 2021; Shu-Chen et al., 2020). The improvement is the ability to fix problems and improve new ideas or products (Nuswowati et al., 2017; OECD, 2021; Vally et al., 2019). The integration between divergent and convergent thinking indicators used to identify individual creative thinking skills has to be adjusted to the learning stages.

Evaluating creative thinking skills in biology covers several stages and dimensions (Runco et al., 2017; Zubaidah et al., 2017). These stages include formulating problems (Runco & Acar, 2012), formulating a hypothesis (Phungsuk et al., 2017), conducting experiments (Nickerson, 2014), and solving problems (Romero et al., 2017; Simper, 2018). The dimension of creative thinking skills can be found in problem-solving by using scientific and social approaches (OECD, 2021; Plucker et al., 2014; Runco et al., 2017), and it can also be found in the ways individuals express themselves by using pictures and written forms (He et al., 2017; Listiana et al., 2016; OECD, 2021; Runco et al., 2017; Watson, 2018). However, the instrument for assessing creative thinking skills which can accommodate divergent and convergent thinking processes in natural science, especially biology, has not been completely formulated. Therefore, this instrument development is necessary to do. Instrument development activities that begin with making a need analysis instrument on how to measure creative thinking skills are currently being carried out.

The assessment of creative thinking skills in science education using the FFOE indicator in divergent thinking processes is still used today. The

measurement of creative thinking skills carried out by Jumadi et al. (2021) used indicators of divergent thinking with a test for high school science students. The instrument was validated using the content validity ratio method. Madyani et al. (2020) measured four indicators of divergent thinking with a test for junior high school science students. The test results used descriptive analysis, showing that the originality is very low. Rudyanto et al. (2019) analyzed the validity and reliability of creative thinking skills using the FFOE indicator through a descriptive analysis of mathematics subjects. So far, the measurement of creative thinking skills only uses divergent thinking and focuses only on students. Meanwhile, the teacher's approval as a facilitator and a evaluator in doing creative assessments is unknown and lack of attention.

The method often used to find out how teachers measure creative thinking skills is through interviews. Matraeva et al. (2020) and As'ari et al. (2019) interviewed teachers about students' creative thinking skills. Interviews can be used to determine the extent to which creative thinking skills have been achieved. However, it is only conducted in a small number of samples. The analysis of the instrument measurement in teacher need analysis activities turns out only to use descriptive analysis, so it does not use valid and reliable testing.

Based on the need for an initial analysis of the development of creative thinking skill instruments, the researcher makes an instrument for teacher needs that integrates divergent and convergent thinking indicators in the steps and dimensions of creative thinking skills. The feasibility of the teacher's tool requirements requires a validation. The instrument by which the teacher needs to evaluate creative thinking skills requires a series of tests to ensure its reliability of the instrument (Baer et al., 2014; Chevalier et al., 2020; Runco & Acar, 2012; Simper, 2018). Thus, the novelty of this study is to analyze the need analysis instrument by making use of the RASCH model.

Accordingly, in this study, the instrument's reliability is analyzed using RASCH (Nielsen, 2018; Sumintono, 2018). The RASCH model is a statistical approach used to measure performance, perception, and attitude (Bonsaksen et al., 2013; Nielsen, 2018). The evaluation of creative thinking skills using the RASCH model has more advantages than classic test theory because it could increase the evaluation quality in quantitative and qualitative studies (Chan et al., 2014). Several advantages of using the RASCH model are: (1) it generates linear and one-dimensional scale; (2) it needs suitability between data and measurement models; (3) it can count error standards; (4) it can estimate the person size as well as the difficulty level of the statement item through the linear scale which is similar to the standard units (logs); and (5) it can check the evaluation system logically and consistently (Planinic et al., 2019).

The RASCH model can analyze the evaluation instrument based on several parameters. For the advantages of the RASCH model, an instrument has to be tested for its reliability, validity, discrimination power, appropriateness, and difficulty level using the RASCH model (Nielsen, 2018; Sumintono & Widhiarso, 2015). The stages are crucial to obtain a reliable evaluation instrument. Therefore, it is necessary to conduct an instrument analysis on the teacher's needs to evaluate creative thinking skills in which the instrument's reliability is analyzed using the RASCH model. Thus, this study aims to analyze the validity, reliability, scale understanding, item difficulties, and bias items of the instrument for analyzing the needs of biology teachers related to the assessment of creative thinking skills by using the RASCH model.

METHOD

This research was conducted as one of the stages of ADDIE research and development. The participants of this research were the 64 Senior High School Biology teachers and the 40 Junior high school Science teachers. The characteristics of the respondents based on age and gender are presented in Table 1. The research was carried out in the Surakarta Residency area, Central Java Province, from July to December 2021. The teacher needs an analysis instrument that validates the teacher's response.

Table 1. Characteristics of Respondents

No	Characteristic of Respondents	Category	Code	Respondents (%)
1	Age category from Depkes RI (2009)	17-25 years old	A	26.92 %
		26-35 years old	B	15.38 %
		36-45 years old	C	13.46 %
		46-55 years old	D	34.61 %
		56-60 years old	E	10.57 %
2	Gender	Male	L	30.76 %
		Female	P	69.23 %

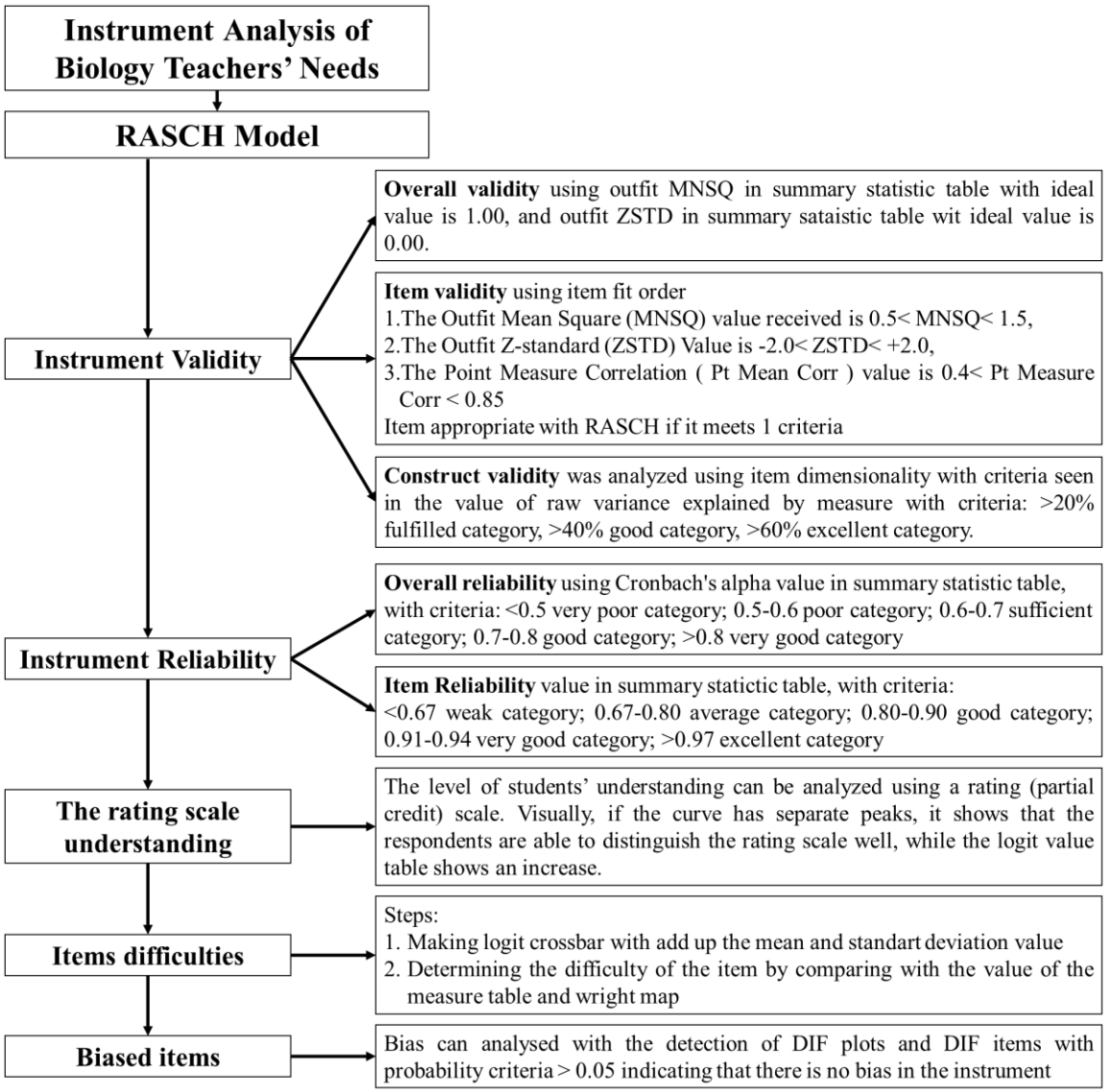


Figure 1. Instrument Analysis Process an Criteria Using RASCH Model (Dahlgren et al., 2017; Pontoppidan et al., 2018; Sumintono & Widhiarso, 2015)

Data collecting technique uses the teacher's needs for an instrument to evaluate the divergent and convergent thinking processes on creative thinking skills in the stages of Biology learning and science learning in schools. The inventory instrument contains statements with a Likert scale of 1 to 4. Data are collected by utilizing the google form application.

Divergent thinking indicators include fluency, flexibility, originality, and elaboration. Meanwhile convergent thinking indicators comprise of fullness, evaluation, and improvement that can be found at the stages of formulating problems, formulating hypotheses, conducting experiments, solving problems, and how students express their learning outcomes as well. The instrument then is analyzed using the RASCH model with the Winstep 3.73.

The analysis stage begins with testing the validity. Validity testing includes overall validity using summary statistics, item validity using item: fit order, and construct validity. The analysis of the instrument's reliability is reviewed using Cronbach's alpha value and the reliability of the items in the statistical summary test. Respondents understand the scale using a partial credit rating scale and a probability curve. Items are analyzed by making a logit ruler which is used to classify item difficulty based on the logit ruler and wright map. The bias items are analyzed using DIF tables and plots.

RESULT AND DISCUSSION

Instrument Validity Analysis

The results of the validity test are divided into two, namely the validity of instrument overall and item (Planinic et al., 2019). The results of the analysis are presented in Table 2.

Table 2. Summary Statistics of the Instrument's Overall Validity Result (Kim, 2021; Sumintono, 2018)

No	Value on Measurement	Measurement Type	Score	Value Category
1	Outfit MNSQ	Item	1.03	Accepted
2	Outfit ZSTD	Item	0.1	Accepted

The results in Table 3 of the instrument validity analysis from the summary statistics show whether the instrument is valid for use or not (Runco & Acar, 2012). Based on the value of the outfit MNSQ item (statement items), the instrument is appropriate to be used for evaluation because the result shows that 1.03 is close to the ideal value of 1.00. Based on the value of the outfit ZSTD item and person, the instrument shows that the data have a logical estimate because the results show that 0.1 is close to the ideal value of 0.00 (Sumintono & Widhiarso, 2015).

Based on AERA & APA, strong validity has the evidence and response validity is the instrument's reliability when the respondents respond. The instrument's validity has been used expert judgment, then directly used to test. Validity testing in the RASCH model informs the quality of the instrument so that validity testing is now more reliable (AERA & APA, 2014). The results of the item dimensionality test can be seen in Table 3 which shows that the construct validity of the instrument has good criteria.

Table 3. The Results of the Item Dimensionality to Analyze Construct Validity

Variance Explained by	Value	Unexpected Variance 1st Contrast of PCA Residuals		Value
Measure	Category	Eigenvalue	Observed	Category
59.1 %	Good	7.4	6.2 %	Good

The results in the unexpected variance 1st contrast of PCA residuals point out good criteria, indicating that all statement items show appropriateness. The result is unidimensionality. It means that the instrument can measure the range of variables or the teachers' responses towards the teacher needs to measure creative thinking skills.

The construct validity of the content variable of the instrument has been able to measure what you want to know. Using the RASCH application model can determine the instrument's construct validity. Based on the research conducted by Madyani et al. (2020), the construct validity has not been analyzed, so this test has a novelty.

The reliability test results on the instrument can be seen from the appropriateness analysis of the statement items used to know which statement item is a misfit. The appropriateness analysis uses the item: fit the order in Table 4. The results show that all statement items can be used to measure the responses (Dahlgren et al., 2017).

Table 4. The Results of the Measurement Item

No	Fulfillment of Criteria	Item Code	Number of Items	Interpretation
1	3	7A, 3A, 4D, 3B, 6A, 3C, 7B, 5D, 1E, 4E, 7D, 4A, 6D, 5B, 1A, 6B, 1B, 3E, 5A, 3D, 7C, 2B	22	Very good
2	2	2G, 5G, 2D, 1C, 5F, 2E, 2C, 4C, 5C, 3F, 6E, 7G, 7E, 7F, 6F, 6G, 3G, 4B, 2A, 6C	20	Good
3	1	1G, 5E, 4F, 4G, 1D, 2F, 1F	7	Can be maintained

Based on Table 4, all items do not require revision to fulfill these criteria. The RASCH analysis model can direct instrument makers to revise items or statements that are not appropriate so that the items have reliability in measurement.

Instrument Reliability Analysis

Table 4 shows the results of the reliability test. The instrument as a whole with alpha Cronbach value 0,91 has a very good category. The reliability of the statement items is 0,99. It has an excellent category (Sumintono & Widhiarso, 2015). The instrument has consistent results if tested in a population (Plucker et al., 2014; Runco & Albert, 1985). The grouping of the statement items has an excellent category because there are 14 categories of the difficulty level of statement items in the instrument. The grouping of the respondents has a good category because there are five levels of respondents' abilities. Thus, the instrument can be used to know the grouping of the statement items and respondents in evaluating creative thinking skills during the learning activities (Göçmen & Coşkun, 2019; Sumintono & Widhiarso, 2015).

Table 5. The Results of the Summary Statistics Measurement to Analyze the Reliability Value and Grouping of Statement Items and Respondents

No	Value on Measurement	Measurement Type	Score	Value Category
1	Alpha Cronbach	Overall	0.91	Very good
2	Reliability	Items	0.99	Excellent

Research conducted by Runco & Acar (2012) states that the reliability value of the creative thinking skills instrument ranges from 0.80 to 0.99. Reliability reviews are often carried out with descriptive analysis and cannot see reliability based on responses and statement items. The RASCH model can distinguish the teacher's responses into groups of item difficulty.

The Rating Scale Understanding Analysis

The evaluation of the rating scale (1, 2, 3, 4) can be seen from the peaks of each scale on the probability curve Figure 2. The image shows separate peaks. Table 6 shows that the value of the logit rating scale has increased from the rating scale (1, 2, 3, 4) with the appropriate difference. The respondents understand the scale well (Sumintono & Widhiarso, 2015).

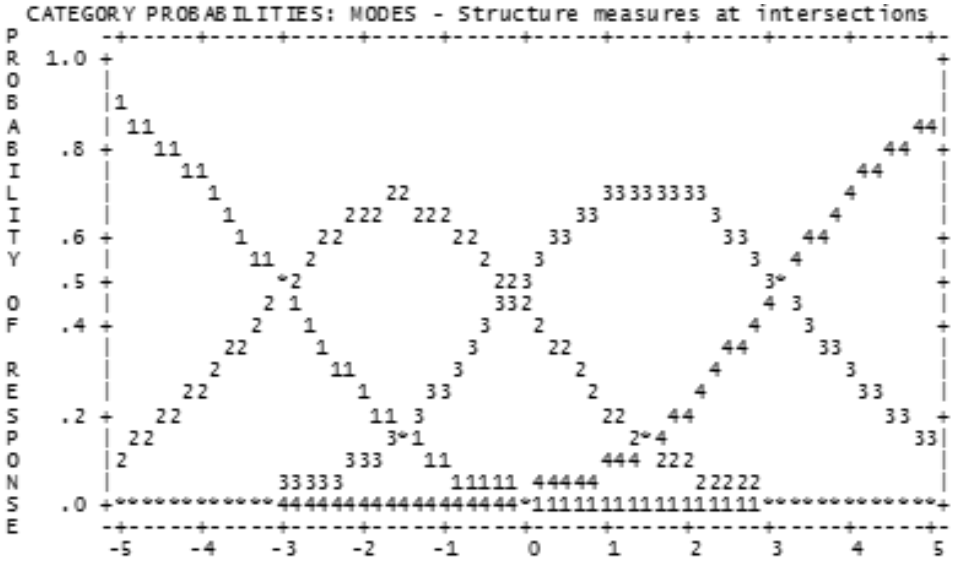


Figure 2. Probability Curve of the Rating Scale Instrument

Table 6. Values on the Rating Scale Measurement to Analyze the Instrument Rating Scale

No	Rating Scale	Value	Andrich Threshold Value
1	1	-4.15	-
2	2	-1.58	-3.01
3	3	1.51	0.13
4	4	4.27	3.14

Respondents' understanding of the scale can be seen through statistics on the RASCH model. The scale is only analyzed through a descriptive analysis when the respondent fills in all the questions. The researcher concludes that the respondent can understand the rating scale well. So far, the understanding of the scale has not been tested. A poor probability curve can be used to analyse the shortcomings of the scale, for example, by reducing the scale range or eliminating a meaningfully neutral rating.

The Difficulty Level of the Statement Items Analysis

The evaluation of the difficulty level of the statement items was carried out using item measures. The statement items' separation or difficulty level was determined by adding the average value with the standard deviation (0.00+1.90= 1.90) used for making the log bar. The log bar can be used as the criteria for determining items that are difficult to approve, approve, and easy to approve in Figure 4. The values of the measure Items and persons need to be considered to know the response given. The results of the person's measures show an average value of (M) -0.2 logit under measure items, which is 0.0 logit. Thus, the teacher's ability to respond is below the average level of difficulty of the standard statement items (Sumintono & Widhiarso, 2015).

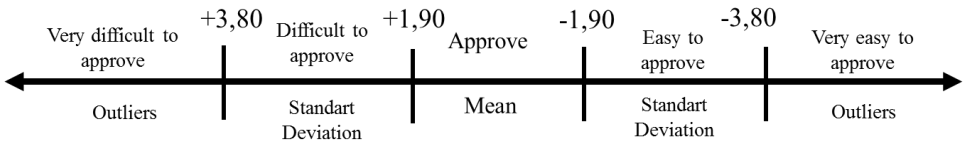
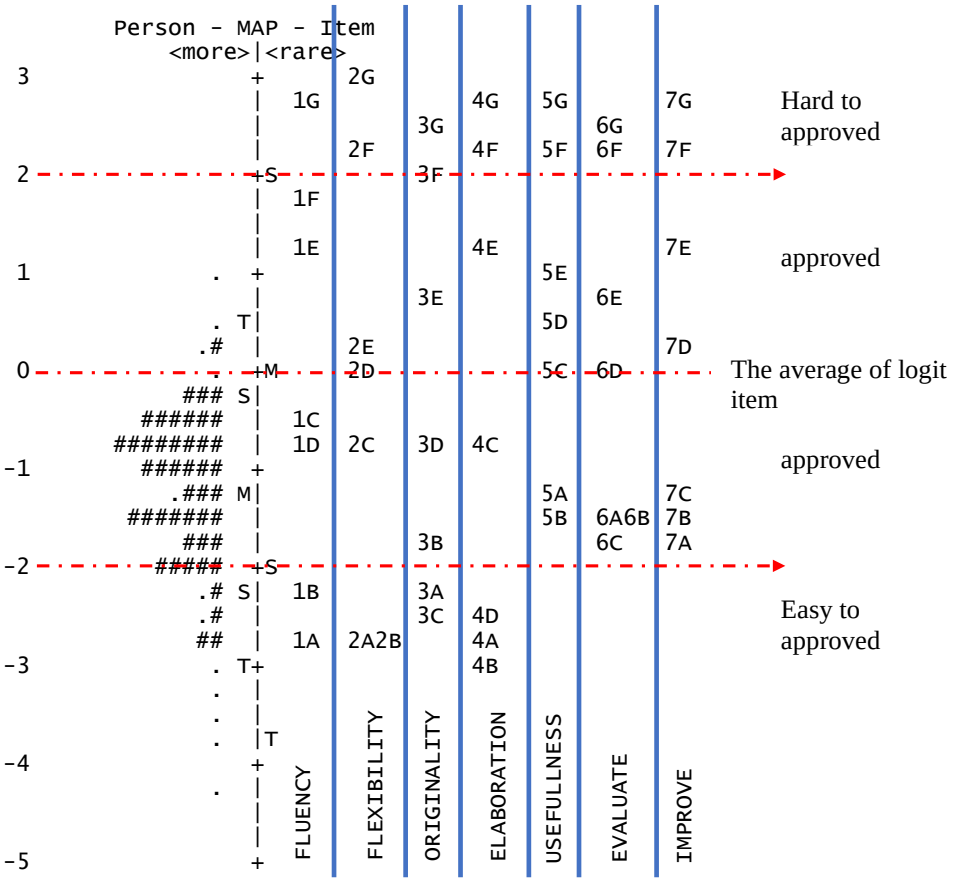


Figure 3. The Difficulty Scale of the Statement Items

The logit bar in Figure 3 is then integrated into the measure items in Table 7 and the wright map in Figure 4 to know the classifications of the statement items. The results show that all teachers do not approve of the evaluation to the dimension of expression method of using pictures and writings. Therefore, the development should accommodate the evaluation of creative thinking skills on ways of expression using pictures and writings (He et al., 2017; Listiana et al., 2016; OECD, 2021). It is stated that biology teachers did not evaluate the creative thinking skills on ways of expressions using pictures and writings because Biology is a natural science. Therefore, teachers more focus on the answers with scientific thinking processes (Rodríguez et al., 2019; Sugiharto et al., 2019; Yustina et al., 2020).



Note : (#) means 2 responses, (.) means 1 response

Figure 4. Wright Item Map to Analyze the Difficulty Level of the Statement Items

Table 7. Item Value: Measure to Know the Difficulty Level of the Statement Items and the Meaning of the Item Code

No	Step Code	Category of Indicator Difficulty and Code is Sorted from the Hardest to the Easiest		
		Difficult to Approve (>1.90 logits)	**Approve (1.90>n>-1.90 logit)**	**Easy to Approve (<-1.90 logit)**
1	A		3 items (5A-7,6A-14.7A-6)	4 items (3A-3.2A-4.1A-4.4A-4)
2	B		4 items (5B-14.6B-14.7B-14.3B-6)	3 items (1B-3,2B-4,4B-1)
3	C		6 items (5C-1,1C-12,2C-16.4C-16.7C-7,6C-6)	1 item (3C-3)
4	D		6 items (5D-1,7D-3,2D-1,6D-1,3D-16,1D-16)	1 item (4D-3)
5	E		7 items (1E-0.4E-0.7E,5E-1.6E-0.3E-0.2E-3)	
6	F	5 items (7F-0.2F-0.5F-0.4F-0.6F-0)	2 items (3F-0.1F-0)	
7	G	7 items (2G-0.1G-0.5G-0.4G-,7G-0.6G-0.3G-0)		

Note: A: Formulating problems; B: Formulating hypothesis; C: Conducting experiments; D: Solving problem with scientific approach; E: Solving problem with social approach; F: expression with writings; G: expression with pictures. 1: Fluency; 2: Flexibility; 3: Originality; 4: Elaboration; 5: Usefulness; 6: Evaluate; 7: Improve.
5A-7 means indicator Usefulness in mentioning ideas with a total of 7 responses

The evaluation of creative thinking skills in problem-solving using social and scientific approaches becomes the dimension not approved by the teachers (Putranta & Supahar, 2019). It is indicated that natural science only focuses on a scientific approach (Chien, 2017; Putranta & Supahar, 2019). However, problem-solving may also implement a social approach, namely the human behavior analysis as well as the effort of the community, government, and social institutions. Therefore, the evaluation instrument development should be able to differentiate between problem-solving using scientific and social approaches (Afacan, 2018). Only a few teachers approve the evaluation of creative thinking skills in the experiment stage, so a special evaluation instrument is needed (Runco & Albert, 1985; Sarriot et al., 2014; Vergara et al., 2018).

The evaluation of creative thinking skills that was carried out was the stages of formulating the hypothesis and problem. This is because in general Biology teachers have already evaluated the hypothesis and problem formulation. Thus, the instrument development should be able to measure the ideas generated and the ability to formulate a hypothesis as well (Sternberg et al., 2020). The research results indicate that few teachers approve the convergent indicators which include usefulness, evaluation, and improvement. Therefore, some items which accommodate the evaluation of the indicators in the evaluation instrument are necessary (Oppezzo & Schwartz, 2014).

The Analysis of Age and Gender Bias towards the Statement Items

The bias test results using DIF in Figure 5 and Table 8 show probability with criteria for bias < 0.05. Five biased items were found, viewed from the age factor, namely 1G (evaluation to fluency) and Item 4G (evaluation to elaboration) on the way of expression using pictures. The difference in the teachers' perception

occurs because evaluation using pictures in Biology learning process is considered ineffective. The older teachers with more than ten years of the working period have many experiences in using evaluation instruments. However, they have not evaluated creative thinking skills on the aspects of fluency and elaboration with pictures and writings (Lia D Rubenstein et al., 2019). Teachers with little experience, whose ages are under 25 years old, tend to give their approval because they have higher motivation to develop learning activities and they are open to the ideas of innovation. In this case, the evaluation of fluency and elaboration is conducted on the way of expression using pictures (Zubaidah et al., 2017).

The evaluation of fluency and elaboration using pictures in Biology learning can be carried out in the class by implementing a learning model with time flexibility such as a blended learning (Sugiharto et al., 2019; Tan, 2009). In the blended learning, the evaluation can accommodate instruments in the form of pictures. Creative thinking evaluation in the aspect of fluency can be realized with an instrument expressed in the form of lines, symbols, or pictures related to the content of Biology (He et al., 2017; Sternberg et al., 2020; Watson, 2018; Zhou et al., 2020). The evaluation of elaboration can be seen from the details or linkages between the illustrated pictures created by the students.

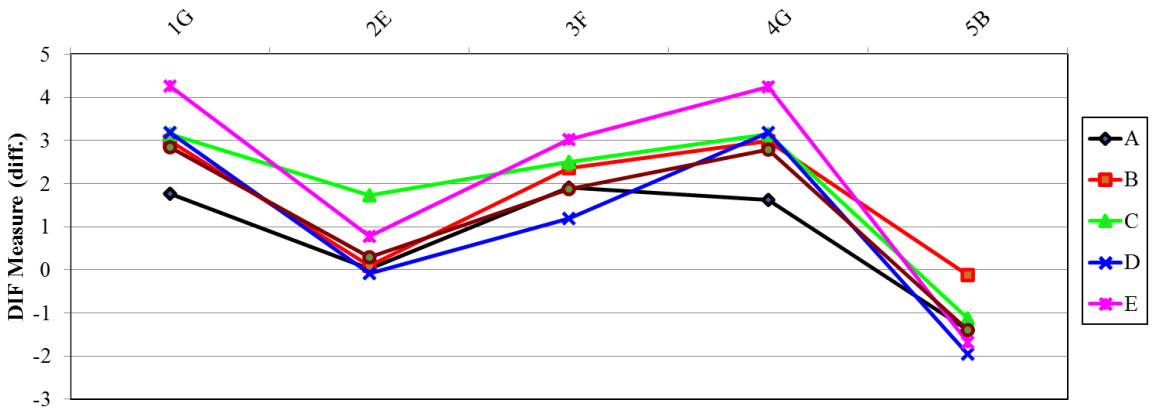


Figure 5. DIF Plot of Age Bias on the Significance of Answering Items

Table 8. Probability Values of Age Differences on the Significance of Answering Items

No	Item Code	Probability Value	Information
1	1G	0.0392	Possibly bias
2	2E	0.0415	Possibly bias
3	3F	0.0490	Possibly bias
4	4G	0.0189	Possibly bias
5	5B	0.0391	Possibly bias

Based on Figure 5 and Table 8, bias on item 2E (Evaluation of flexibility) on problem-solving using a social approach occurs because late-middle-aged and old-aged teachers disapproved. However, adult-aged teachers and middle-aged teachers approved. In this sense, the social environment factors are very influential towards the diversity of the teacher's point of view (Boelens et al., 2018; Lisa D Rubenstein et al., 2018). The teachers who interact actively with society have better understanding and opportunity to identify behaviors causing problems such as the environmental pollution, habitat destruction, and loss of biodiversity (Huang et al., 2019). The teachers should be able to make innovations towards social issues related to the evaluation of flexibility using various social approaches (Boelens et al., 2018). Different perceptions may occur due to the different

social interactions among different age of teachers (Lisa D Rubenstein et al., 2018).

Bias on item 3F (Evaluation of originality) on the way of writing expression occurs because elderly teachers, early adult teachers, and late adult teachers do not approve. However, the late-teenager teachers and the early-old teachers approved. These results may have occurred because of the diversity of the teachers' ways of evaluating students' writings. One of the factors influencing the evaluation of the students' writings is teachers' subjectivity. Some teachers evaluate the appropriateness of the report's title without considering the novelty or the originality of the student's writing. For example, sometimes teachers give maximum scores to intelligent and diligent students without checking the originality or novelty of the student's ideas (Gralewski & Karwowski, 2019).

Furthermore, teachers' experiences become the other factors influencing the evaluation of the student's writing because the teachers with better experience are more familiar with the student's writing so that they can check the students' writing better (Schoevers et al., 2019; Wang et al., 2018; Watson, 2018). The evaluation of originality in writing can be done by checking the originality of a student's writing and the novelty of the student's ideas (Kafipour et al., 2018). Some methods can be used to express ideas in writing, such as explanation texts. Explanation text writing tasks enable teachers to identify the originality of students' writing and ideas due to explanation text written based on the construction of students' knowledge (Göçmen & Coşkun, 2019). In this respect, Turnitin can be used as a means of checking the originality of the student's writing if the teachers' experience has not yet been sufficient (Matheson & Starr, 2013).

Bias on item 5B (evaluation of usefulness) on formulating hypothesis occurs because late-teenager teachers only approve, while the other age-range teachers are easy to approve (Sumintono & Widhiarso, 2015). These indicate that teachers of all ages have paid attention to acknowledge the usefulness of the hypothesis formulated by the students. Usefulness emphasizes on the ability of individuals to mention the purpose of hypothesis testing (Charyton et al., 2011). Biology learning is natural science knowledge which emphasizes on scientific steps, so that hypothesis formulation requires creative thinking skills. Contextual learning can be used to give opportunity for the students to formulate hypotheses based on the real problems.

Elaboration on problem-solving with a social approach can be seen from the order of the explanation of the problem solving analyzed from a social perspective, for instance, the correlation between the behavior of the causes of problems, social analysis towards a case, and the selection of problem-solving (Bart et al., 2017). Based on Figure 5 and Table 9, four items are potentially gender-biased. This is indicated by the probability value below the significance value< 0.05. Item 4D shows that women can easily approve of the elaboration evaluation on problem solving with a social approach. This difference of perception may occur because female teachers tend to have higher motivation in socialization than male teachers (Webb & Rule, 2014)

Table 9. Probability Values of Gender Differences on the Significance of Answering Items

No	Item code	Probability Value	Description
1	4D	0.0251	Probably bias
2	6A	0.0357	Probably bias

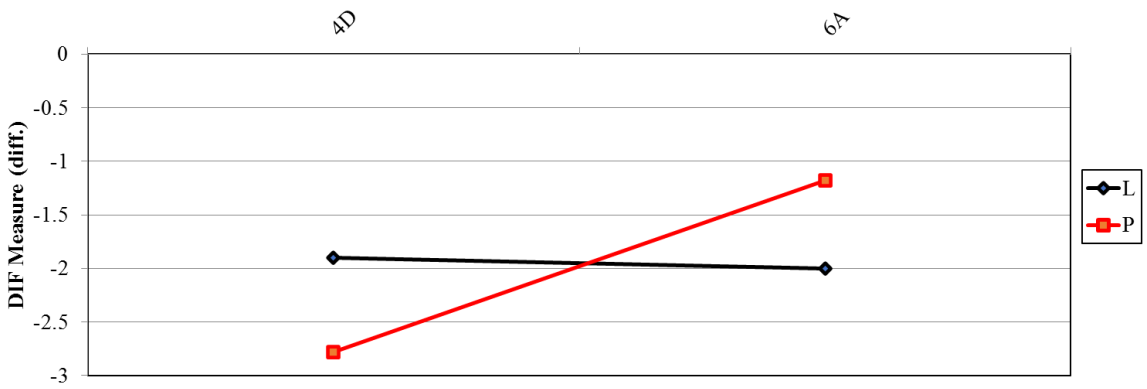


Figure 6. DIF Plot of Gender Bias on the Significance of Answering Items

Percentage Based on Table 9 and Figure 6, bias in item 6A occurs because men are easier to approve than women. This difference of perception between male teachers and female teachers may occur due to several factors such as motivation for socialization, teaching motivation, and psychology which are different between male and female teachers. Thus, these differences cause a difference in male and female teachers' teaching and evaluation styles (Wu et al., 2019). Evaluation assessment has been done in problem formulation (Zubaidah et al., 2017). The purpose of evaluating the students in formulating new problems is to foster the students' comprehension (Shu-Chen et al., 2020). As such, the teachers should give the special assignments to measure how the students evaluate their ideas (Bedir, 2019).

CONCLUSION

Based on the results of this research, it can be concluded that the teacher response instrument has good criteria to be applied to know the teachers' approval to have conducted evaluations on students' creative thinking skills in scientific learning stages and creative thinking dimension. Based on the analysis, the overall validity is acceptable, and the item validity does not require improvement. Based on the analysis, the overall reliability is very good, and the item reliability is excellent. Rating scale analysis shows that respondents have understood the Linkert scale of 1 to 4 well. Then, based on the item difficulties results of the teachers' responses, it is found that there have not been any teachers in evaluating the indicators of creative thinking skills through students' expressions using writing or pictures in problem-solving assessment by using scientific and social approaches. The bias test results on the instrument items indicate that five items could be biased due to age differences and two could be biased due to gender types. Thus, the development of assessment instruments to measure creative thinking skills with scientific and social problem-solving based assignments as well as writing and visual expressions, is definitely needed.

REFERENCES

Afacan, Y. (2018). Student experiences of blended learning in interior architecture. Journal of Information Technology Education: Research, 17(1), 399–422. https://doi.org/10.28945/4122

As'ari, A. R., Kurniati, D., & Subanji, S. (2019). Teachers expectation of students' thinking processes in written works: A survey of teachers' readiness in making thinking visible. Journal on Mathematics Education, 10(3), 409–424. https://doi.org/10.22342/jme.10.3.7978.409-424

Baer, J., Kaufman, J. C., Gentile, C. A., Baer, J., Kaufman, J. C., & Gentile,

C. A. (2014). Extension of the consensual assessment technique to nonparallel creative products. Creativity Research Journal, 16(1), 113–117. https://doi.org/10.1207/s15326934crj1601_11

Barbot, B., & Lubart, T. (2012). Creative thinking in music: Its nature and assessment through musical exploratory behaviors. Psychology of Aesthetics, Creativity, and the Arts, 6(3), 231–242. https://doi.org/10.1037/a0027307

Bart, W. M., Hokanson, B., & Can, I. (2017). An investigation of the factor structure of the torrance tests of creative thinking. Educational Sciences: Theory & Practice, 17(2), 515–528. https://doi.org/10.12738/estp.2017.2.0051

Bedir, H. (2019). Pre-service ELT teachers' beliefs and perceptions on 21st century learning and innovation skills (4Cs). Journal of Language and Linguistic Studies, 15(1), 231–246. https://doi.org/10.17263/jlls.547718

Benedek, M., Fink, A., & Neubauer, A. C. (2006). Enhancement of ideational fluency by means of computer-based training. Creativity Research Journal, 18(3), 317–328. https://doi.org/10.1207/s15326934crj1803_7

Boelens, R., Voet, M., & De Wever, B. (2018). The design of blended learning in response to student diversity in higher education: Instructors' views and use of differentiated instruction in blended learning. Computers and Education, 120, 197–212. https://doi.org/10.1016/j.compedu.2018.02.009

Bonsaksen, T., Kottorp, A., Gay, C., Fagermoen, M., & Lerdal, A. (2013). Rasch analysis of the general self-efficacy scale in a sample of persons with morbid obesity. Health and Quality of Life Outcomes, 11(1), 202. https://doi.org/10.1186/1477-7525-11-202

Catarino, P., Vasco, P., Lopes, J., Silva, H., & Morais, E. (2019). Cooperative learning on promoting creative thinking and mathematical creativity in higher education. Revista Iberoamericana Sobre Calidad, Eficacia y Cambio En Educacion, 17(3), 5–22. https://doi.org/10.15366/reice2019.17.3.001

Chan, S. W., Ismail, Z., & Sumintono, B. (2014). A rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics. Procedia - Social and Behavioral Sciences, 129, 133–139. https://doi.org/https://doi.org/10.1016/j.sbspro.2014.03.658

Charyton, C., Jagacinski, R. J., Merrill, J. A., Clifton, W., & Dedios, S. (2011). Assessing creativity specific to engineering with the revised creative engineering design assessment. Journal of Engineering Education, 100(4), 778–799. https://doi.org/10.1002/j.2168-9830.2011.tb00036.x

Chevalier, M., Giang, C., Piatti, A., & Mondada, F. (2020). Fostering computational thinking through educational robotics: A model for creative computational problem solving. International Journal of STEM Education, 7(1), 39. https://doi.org/10.1186/s40594-020-00238-z

Chien, H.-M. (2017). Study of volunters' science service satisfaction in relation to their self-directed. Journal of Baltic Science Education, 16(2), 188–198. https://doi.org/10.33225/jbse/17.16.188

Dahlgren, A. A., Guttersrud, Ø., Nsangi, A., & Semakula, D. (2017). Measuring ability to assess claims about treatment effects: A latent trait

analysis of items from the 'Claim Evaluation Tools' database using Rasch modelling. BMJ Open, 1(3), 1–10. https://doi.org/10.1136/bmjopen-2016-013185

Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (2020). Global innovation index 2020 who will finance innovation ? In Global Innovation Index 2020 (p. 398). Cornell University, INSEAD, and the World Intellectual Property Organization.

Göçmen, Ö., & Coşkun, H. (2019). The effects of the six thinking hats and speed on creativity in brainstorming. Thinking Skills and Creativity, 31, 284–295. https://doi.org/10.1016/j.tsc.2019.02.006

Gralewski, J., & Karwowski, M. (2019). Are teachers' ratings of students' creativity related to students' divergent thinking? A meta-analysis. Thinking Skills and Creativity, 33. https://doi.org/10.1016/j.tsc.2019.100583

Guilford, J. P. (1975). Varieties of creative giftedness, their measurement and development. Gifted Child Quarterly, 19(2), 107–121. https://doi.org/10.1177/001698627501900216

Hass, R. W. (2015). Feasibility of online divergent thinking assessment. Computers in Human Behavior, 46, 85–93. https://doi.org/10.1016/j.chb.2014.12.056

He, W. jing, Wong, W. chi, & Chan, M. ki. (2017). Overexcitabilities as important psychological attributes of creativity: A Dabrowskian perspective. Thinking Skills and Creativity, 25, 27–35. https://doi.org/10.1016/j.tsc.2017.06.006

Huang, X., Chi-Kin Lee, J., & Yang, X. (2019). What really counts? Investigating the effects of creative role identity and self-efficacy on teachers' attitudes towards the implementation of teaching for creativity. Teaching and Teacher Education, 84, 57–65. https://doi.org/10.1016/j.tate.2019.04.017

Jasim Mohammed, H., & Ati Daham, H. (2021). Analytic hierarchy process for evaluating flipped classroom learning. Computers, Materials & Continua, 66(3), 2229–2239. https://doi.org/10.32604/cmc.2021.014445

Jia, X., Hu, W., Cai, F., Wang, H., Li, J., Runco, M. A., & Chen, Y. (2017). The influence of teaching methods on creative problem finding. Thinking Skills and Creativity, 24, 86–94. https://doi.org/10.1016/j.tsc.2017.02.006

Jumadi, J., Perdana, R., Hariadi, M. H., Warsono, W., & Wahyudi, A. (2021). The impact of collaborative model assisted by google classroom to improve students' creative thinking skills. International Journal of Evaluation and Research in Educaion, 10(2), 396–403. https://doi.org/10.11591/ijere.v10i2.20987

Kafipour, R., Mahmoudi, E., & Khojasteh, L. (2018). The effect of task-based language teaching on analytic writing in EFL classrooms. Cogent Education, 5(1), 1496627. https://doi.org/10.1080/2331186X.2018.1496627

Kim, J. (2021). Development and validation of the career adaptability scale for undergraduates in Korea. Sustainability, 13(11004), 2–25. https://doi.org/. https://doi.org/10.3390/su131911004

Listiana, L., Susilo, H., & Suwono, H. (2016). Empowering students' metacognitive skills through new teaching strategy (group

investigation integrated with think talk write) in biology classroom. Journal of Baltic Science Education, 15(3), 391–400. https://doi.org/https://doi.org/10.33225/jbse/16.15.391

Lucas, B. (2016). A five-dimensional model of creativity and its assessment in schools. Applied Measurement in Education, 29(4), 278–290. https://doi.org/10.1080/08957347.2016.1209206

Madyani, I., Yamtinah, S., Utomo, S. B., Saputro, S., & Mahardiani, L. (2020). Profile of students' creative thinking skills in science learning. Proceedings of the 3rd International Conference on Learning Innovation and Quality Education (ICLIQE 2019), 397. https://doi.org/10.2991/assehr.k.200129.119

Matheson, L. G., & Starr, S. (2013). Is it cheating or learning the craft of writing? Using turnitin to help students avoid plagiarism. Research in Learning Technology, 21(1063519), 1–13. https://doi.org/10.3402/rlt.v21i0.17218

Matraeva, A. D., Rybakova, M. V, Vinichenko, M. V, & Oseev, A. A. (2020). Development of creativity of students in higher educational institutions: Assessment of students and expert. Universal Journal of Educational Research, 8(1), 8–16. https://doi.org/10.13189/ujer.2020.080102

Nickerson, R. S. (2014). Enhancing creativity. In Handbook of Creativity (pp. 392–430). Cambridge University Press. https://doi.org/10.1017/cbo9780511807916.022

Nielsen, T. (2018). The intrinsic and extrinsic motivation subscales of the motivated strategies for learning questionnaire: A Rasch-based construct validity study. Cogent Education, 5(1), 1504485. https://doi.org/10.1080/2331186X.2018.1504485

Nuswowati, M., Susilaningsih, E., Ramlawati, & Kadarwati, S. (2017). Implementation of problem-based learning with green chemistry vision to improve creative thinking skill and students' creative actions. Jurnal Pendidikan IPA Indonesia, 6(2), 221–228. https://doi.org/10.15294/jpii.v6i2.9467

OECD. (2021). PISA 2021 creative thinking framework (third draft). Oecd, 53(9), 1689–1699.

Oppezzo, M., & Schwartz, D. L. (2014). Give your ideas some legs: The positive effect of walking on creative thinking. Journal of Experimental Psychology, 40(4), 1142–1152. https://doi.org/https://doi.org/10.1037/a0036577

Phungsuk, R., Viriyavejakul, C., & Ratanaolarn, T. (2017). Development of a problem-based learning model via a virtual learning environment. Kasetsart Journal of Social Sciences, 38(3), 297–306. https://doi.org/10.1016/j.kjss.2017.01.001

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. Physical Review Physics Education Research, 15(2), 20111.

Plucker, J. A., Qian, M., & Schmalensee, S. L. (2014). Is what you see what you really get? Comparison of scoring techniques in the assessment of real-world divergent thinking. Creativity Research Journal, 26(2), 135–143. https://doi.org/10.1080/10400419.2014.901023

Pontoppidan, M., Nielsen, T., & Kristensen, I. H. (2018). Psychometric properties of the

danish parental stress scale: Rasch analysis in a sample of mothers with infants. PLoS ONE, 13(11), 1–20. https://doi.org/https://doi.org/10.1371/journal. pone.0205662

Putranta, H., & Supahar. (2019). Development of physics-tier tests (PysTT) to measure students' conceptual understanding and creative thinking skills: a qualitative synthesis. Journal for the Education of Gifted Young Scientists, 7(3), 747–775. https://doi.org/10.17478/jegys.587203

Rodríguez, G., Pérez, N., Núñez, G., Baños, J.-E., & Carrió, M. (2019). Developing creative and research skills through an open and interprofessional inquiry-based learning course. BMC Medical Education, 19(1), 134. https://doi.org/10.1186/s12909-019-1563-5

Romero, M., Lepage, A., & Lille, B. (2017). Computational thinking development through creative programming in higher education. International Journal of Educational Technology in Higher Education, 14(1), 1–15. https://doi.org/10.1186/s41239-017-0080-z

Rubenstein, Lia D, Callan, G. L., Ridgley, L. M., & Henderson, A. (2019). Students' strategic planning and strategy use during creative problem solving: The importance of perspective-taking. Thinking Skills and Creativity, 34, 100556. https://doi.org/10.1016/j.tsc.2019.02.004

Rubenstein, Lisa D, Ridgley, L. M., Callan, G. L., Karami, S., & Ehlinger, J. (2018). How teachers perceive factors that influence creativity development: Applying a social cognitive theory perspective. Teaching and Teacher Education, 70, 100–110. https://doi.org/10.1016/j.tate.2017.11.012

Rudyanto, H. E., Ghufron, A., & Hartono. (2019). Use of integrated mobile application with realistic mathematics education: A study to develop elementary students' creative thinking ability. International Journal of Interactive Mobile Technologies, 13(10), 19–27. https://doi.org/10.3991/ijim.v13i10.11598

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. Creativity Research Journal, 24(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., Acar, S., & Cayirdag, N. (2017). A closer look at the creativity gap and why students are less creative at school than outside of school. Thinking Skills and Creativity, 24, 242–249. https://doi.org/10.1016/j.tsc.2017.04.003

Runco, M. A., & Albert, R. S. (1985). The reliability and validity of ideational originality in the divergent thinking of academically gifted and nongifted children. Educational and Psychological Measurement, 45(3), 483–501. https://doi.org/10.1177/001316448504500306

Sarriot, E. G., Kouletio, M., Jahan, S., Rasul, I., & Musha, A. (2014). Advancing the application of systems thinking in health: Sustainability evaluation as learning and sense-making in a complex urban health system in Northern Bangladesh. Health Research Policy and Systems, 12(1), 45. https://doi.org/10.1186/1478-4505-12-45

Schoevers, E. M., Leseman, P. P. M., Slot, E. M., Bakker, A., Keijzer, R., & Kroesbergen, E. H. (2019). Promoting pupils' creative thinking

in primary school mathematics: A case study. Thinking Skills and Creativity, 31, 323–334. https://doi.org/10.1016/j.tsc.2019.02.003

Shu-Chen, C., Gwo-Jen, H., & Chiu-lin, L. (2020). Effects of the group leadership promotion approach on students' higher order thinking awareness and online interactive behavioral patterns in a blended learning environment. Interactive Learning Environments, 28(2), 246–263. https://doi.org/10.1080/10494820.2019.1636075

Simper, N. (2018). Rubric authoring tool supporting cognitive skills assessment across an institution. Teaching and Learning Inquiry, 6(1), 10–24. https://doi.org/10.20343/teachlearninqu.6.1.3

Sternberg, R. J., Todhunter, R. J. E., Litvak, A., & Sternberg, K. (2020). The relation of scientific creativity and evaluation of scientific impact to scientific reasoning and general intelligence. Journal of Intelligence, 8(2), 17. https://doi.org/10.3390/jintelligence8020017

Sugiharto, B., Corebima, A. D., Susilo, H., & Ibrohim. (2019). The pre-service biology teacher readiness in Blended Collaborative Problem Based Learning (BCPBL). International Journal of Instruction, 12(4), 113–130. https://doi.org/10.29333/iji.2019.1248a

Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. Advances in Social Science, Education and Humanities Research, 173(Icei 2017), 38–42. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2015). Aplikasi pemodelan rasch pada assessment pendidikan. Trim komunikata.

Tan, O. (2009). Problem-based learning and creativity. Cengage Learning Asia Pte Ltd.

Tran, T. B. L., Ho, T. N., Mackenzie, S. V., & Le, L. K. (2017). Developing assessment criteria of a lesson for creativity to promote teaching for creativity. Thinking Skills and Creativity, 25, 10–26. https://doi.org/10.1016/j.tsc.2017.05.006

Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times (J. Wiley (ed.); First Edit). Jossey-Bass.

Vally, Z., Salloum, L., AlQedra, D., El Shazly, S., Albloshi, M., Alsheraifi, S., & Alkaabi, A. (2019). Examining the effects of creativity training on creative production, creative self-efficacy, and neuro-executive functioning. Thinking Skills and Creativity, 31, 70–78. https://doi.org/10.1016/j.tsc.2018.11.003

Vergara, M. C., Barrios Galleguillos, N., Jofré Cuello, L., Alvarez-Marin, A., & Acuña-Opazo, C. (2018). Does socioeconomic status influence student creativity? Thinking Skills and Creativity, 29, 142–152. https://doi.org/10.1016/j.tsc.2018.07.005

Wang, Q., Huang, C., & Quek, C. L. (2018). Students' perspectives on the design and implementation of a blended synchronous learning environment. Australasian Journal of Educational Technology, 34(1), 1. https://doi.org/https://doi.org/10.14742/ajet.3404

Watson, J. (2018). Deferred creativity: Exploring the impact of an undergraduate learning experience on professional practice. Teaching and Teacher Education, 71, 206–213. https://doi.org/10.1016/j.tate.2017.12.018

Webb, A. N., & Rule, A. C. (2014). Effects of teacher lesson introduction on second graders' creativity in a science/literacy integrated unit on health and nutrition. Early Childhood Education Journal, 42(5), 351–360. https://doi.org/10.1007/s10643-013-0615-4

Wu, Y., Lian, K., Hong, P., Liu, S., Lin, R.-M., & Lian, R. (2019). Teachers' emotional intelligence and self-efficacy: Mediating role of teaching performance. Social Behavior and Personality: An International Journal, 47(3), 1–10. https://doi.org/10.2224/sbp.7869

Yustina, Y., Syafii, W., & Vebrianto, R. (2020). The effects of blended learning and project-based learning on pre-service biology teachers' creative thinking skills through online learning in the COVID-19 pandemic. Jurnal Pendidikan IPA Indonesia, 9(3), 408–420. https://doi.org/10.15294/jpii.v9i3.24706

Zhou, N., Kisselburgh, L., Chandrasegaran, S., Badam, S. K., Elmqvist, N., & Ramani, K. (2020). Using social interaction trace data and context to predict collaboration quality and creative fluency in collaborative design learning environments. International Journal of Human-Computer Studies, 136, 102378. https://doi.org/10.1016/j.ijhcs.2019.102378

Zubaidah, S., Fuad, N. M., Mahanal, S., & Suarsini, E. (2017). Improving creative thinking skills of students through differentiated science inquiry integrated with mind map. Journal of Turkish Science Education, 14(4), 77–91. https://doi.org/10.12973/tused.10214a