Creativity studies
ISSN 2345-0479 / eISSN 2345-0487
2022 Volume 15 Issue 2: 560–576
https://doi.org/10.3846/cs.2022.15182

THE ANALYSIS OF CONSTRUCT VALIDITY OF INDONESIAN CREATIVITY SCALE USING RASCH MODEL

Dian INDIHADI [1, *], Dodi SURYANA [2], Aslina Binti AHMAD [3]

[1]Department of Elementary Teacher Education, Indonesia University of Education, Jl. Dadaha No. 18, Tasikmalaya, 46115 Bandung, Indonesia
[2]Department of Guidance and Counseling, Faculty of Educational Science, Indonesia University of Education, Jl. Dr. Setiabudhi No. 229, 40154 Bandung, Indonesia
[3]Department of Psychology and Counseling, Faculty of Human Development, Sultan Idris Education University, Tanjung Malim Perak, 35900 Tanjung Malim, Malaysia

Received 22 June 2021; accepted 1 March 2022

Abstract. This study was conducted to analyze the construct validity of the Indonesian creativity scale developed by the Integrated Services Unit for Guidance and Counseling Services at the Indonesia University of Education, Indonesia. The research sample was 500 elementary school students in Indonesia, with 22 samples in West Java, Indonesia. This study employed a quantitative approach and a descriptive method with a survey design as the research method. Data were analyzed using the Rasch model using *Winstep* version 5.1.4. The analysis results show that the interaction between respondents and items is included in the very good category. The respondent's reliability value which shows the consistency of the respondent in filling out the instrument is included in the very good category. The reliability of the instrument shows that it is in the very good category. However, the unidimensionality of the instrument indicates that the instrument is still not perfect for measurement. In addition, four items do not meet the measurement standards, which have a logit value that is greater than the average value after being added to the standard deviation value, and some item which has a bias in the gender category. It is hoped that with the results of this study, the Indonesian creativity scale instrument can be evaluated and developed so that an accurate instrument is obtained in measuring the creativity of elementary school students.

Keywords: creativity, developmentally appropriate practice, elementary student, Indonesian creativity scale, Rasch model.

Introduction

Creativity is an ability that is fundamental for all individuals. Creativity provides many updates that are very helpful for individual life. Creative ideas emerge and even have a better impact on any sector (Chakrabarti & Bligh, 1994; Kowsar & Mukherjee, 2021; Quiñones-Gómez, 2021; Shah et al., 2003), and this is a tangible form that creativity is an ability that

*Corresponding author. E-mail: *dianindihadi@upi.edu*

every individual must-have. In other words, individuals will continue to think about how to create something that can have a good influence on the quality of their lives, so creativity is the reason why ideas keep popping up in every aspect of life (Long & Henderson, 1965; Zosh et al., 2017).

In the context of elementary school students, creativity becomes a critical ability to be measured and then developed. Several studies (Braslauskas, 2021; Du et al., 2019; Wilson, 2007) reveal that creativity needs to be instilled from an early age to provide the ability to see the problems that arise in their lives creatively for now and in the future. It is also mentioned that intelligence in children, which subsequently even affects their achievement, cannot be separated from creativity (Fuchs-Beauchamp et al., 1993; Glăveanu, 2010; Sen & Hagtvet, 1993; Sternberg, 2003). Children with high creativity can be said to have a good level of intelligence. However, in the process of developing creativity, it is necessary to pay attention to how the level of creativity in children is, what aspects – in creativity – need to be fulfilled by children, so that the development of creative abilities will be adequately served because they pay attention to the position of their creative development. Therefore, an accurate measuring tool is needed to ensure the position of creativity development in children.

The Indonesian creativity scale (ICS) is an instrument that measures creativity in children, especially those aged 9 to 12 years. ICS was developed by the Integrated Services Unit for Guidance and Counseling Services (ISUGCS) at the Indonesia University of Education (IUE). The instrument adopts the creativity instrument developed by Guilford (1959), which includes aspects of fluency, flexibility, and originality.

ICS has been tested and has good validity. However, the validation process still uses classical test theory (CTT). The effectiveness of CTT has limitations, one of which is that when two different types of tests are given to two different groups, the results obtained cannot be compared. In addition, the level of difficulty and item weighting in the CTT is highly dependent on the measurements carried out.

The analysis of the Rasch model (RM), which is an item response theory (IRT) model developed by Georg Rasch around 1960, is here to provide a solution to the shortcomings of CTT (Higgins, 2007). The RM analysis accommodates a probability approach in viewing a measurement object so that the analysis of the RM is not deterministic and can identify the measured object more accurately. In addition, Rasch modeling can overcome metric differences between items and overcome the problem of data intervals (Andrich, 1988).

Therefore, this study was conducted to analyze ICS using the RM to provide information about the quality of the instrument based on the IRT model in measuring creativity. Thus, the instrument can be evaluated to get the best measuring tool to determine the development of creativity in children.

1. Research context

Indonesia is a developing country with more than 250 million people from various tribes, races, and ethnicities. It is very much in need of much innovation from its people to encourage development in all life sectors. In the Global Innovation Index in 2020, Indonesia is still very far from becoming an innovative country. Indonesia is in the same rank as the previous year, which is ranked 85 out of 131 countries, and with this ranking, Indonesia is in the lower middle category in terms of innovation (Dutta et al., 2020). This is a big challenge for all Indonesian people, especially people who focus on primary education to develop community creativity to create innovative ideas.

The education sector is a very strategic place to develop creativity. In the educational process, student development is continuously monitored to reach maturity. In other words, the educational process is carried out based on student needs (Departemen Pendidikan Nasional, 2008). Creativity, which is a significant capital for individuals to deal with all problems in the development process (Chiu et al., 2019; Yamaoka & Yukawa, 2020), is one of the essential needs of the students. Guidance-based learning certainly requires a basis in the process, that is, the results of the assessment. However, it will not be accurate if the measuring instrument used does not have good validity.

Based on these problems, finding valid instruments to determine student development, especially the development of student creativity, is essential to note. The guidance-based learning process will not be optimal if the measuring instrument used to determine student development is not valid. Therefore, testing the validity of the instruments to determine the development of students' creativity is vital to help students reach maturity in their development. Thus, Indonesia will have creative human resources to create innovations to help Indonesia become an innovative country.

2. Method

The study was conducted using a descriptive method with a quantitative approach and a cross-sectional research design. The research subjects were elementary school students in West Java; those in Cirebon, West Java; Tasikmalaya, West Jawa; and Sukabumi, West Java, were randomly selected using a random sampling technique.

2.1. Rasch model

The RM is one of the most popular IRT models. The general framework of mathematical functions specifically describes the interaction between persons and items. IRT does not depend on a sample of certain items or people selected in an exam so that measurements made using this model are more precise and items can be calibrated (Andrich, 2010; Nurhudaya et al., 2019).

The RM as one of the IRT models has the advantage that it can provide accurate information when testing instruments. The RM accommodates a probability approach in viewing the attributes of the measured object. This makes the RM deterministic so that it can identify the measuring object more accurately. In addition, the raw score is not a final measurement result but is still temporary. There is an error, namely the difference in metrics between items that can have an impact on a person's level of ability to answer. The RM can overcome these differences in metrics between items by calibrating so that it can place items and subjects in the same matrix. That way, the resulting score is not a raw score but a pure score. The RM also applies logarithms to the odd ratio function. This can overcome the problem of data intervals, which if you pay attention to the interval data used, such as a scale of 1 to 4, it does not only have a distance of 1 for each scale because this is based on the opinion given by the subject. In addition, the RM has met objective measurements. This is because the RM uses estimation and calibration techniques to produce data that is free from the influence of the type of subject, the characteristics of the rater, and the characteristics of the measuring instrument (Mohamad et al., 2015; Engelhard & Wind, 2018; Yasin et al., 2015; Bond & Fox, 2015; Boone et al., 2014).

3. Participant

The research participants were elementary school students in West Java, specifically in Cirebon, Tasikmalaya, and Sukabumi, who were randomly selected. West Java is a province used as a place to research because it is unique compared to other provinces. The majority of the population in West Java is Sundanese, and it is a province that has areas with contrasting characteristics of two identities where there are urban communities and traditional communities. Along with technological developments and economic developments, large population movements have occurred in West Java until this year. This makes the population in West Java is not only derived from indigenous people but also other tribes and ethnicities.

Research participants were selected from three different cities in West Java. These cities were chosen based on the results of the consultations carried out with geography experts to ensure that the data taken could represent the province of West Java. Cirebon is an area located on Java island's, Indonesia, coast or the "pantura" route that connects Jakarta, Indonesia–Semarang, Central Java, Indonesia–Cirebon–Surabaya, East Java, Indonesia. In addition, Tasikmalaya is located in the South-Eastern part of West Java, which is a city district that has the largest area in West Java. The last is Sukabumi, which is located in the South-West of West Java province. Participants were selected using a random sampling technique in several schools. Table 1 described the research participants.

The data collected with a creativity scale instrument had been previously notified to the school with the information that it would be used for research purposes only, and the data privacy is safe.

Table 1. Research participant (source: created by authors)

Region in West Java, Indonesia	Gender	Amount
Cirebon	Male	4
	Female	4
Tasikmalaya	Male	3
	Female	4
Sukabumi	Male	3
	Female	4
Total		22

4. Indonesian creativity scale

The ICS is an instrument used to measure the creativity of individuals who are at the age of 9–12 years or equal to grade 4–6 of an elementary school in Indonesia's education system. The instrument was developed by the ISUGCS of IUE. The instrument was developed based on the theory proposed by Guilford in the form of a rating scale.

Guilford (1950) outlines that there are several factors used to determine individual creativity, such as:

- The first is fluency. Creativity in individuals will be seen from how many ideas are created in a limited time. For example, the individual will be given an object in a flat shape, and the individual is instructed to make something from the provided flat shape. The more individuals make drawings from the provided flat shapes, the more individuals can create ideas quickly with a limited time (Guilford, 1962). In simple terms, fluency can be interpreted as how easy and fluid individuals are in creating ideas;
- The second is flexibility. An idea is created as a result of individual thinking, but it should be noted that an idea is a solution to a problem. Based on this statement, it is known that ideas are created when a problem occurs in an individual. However, it should be understood that there is not only one type of problem in the life of an individual. There are many types of problems with different ways to solve them. Therefore, ideas that are solutions derived from the thinking process to solve problems should not only be made for one problem but also various problems (de Dreu et al., 2011; Sisk, 2014). In other words, individuals also need to have sufficient flexibility to process ideas by looking at the types of problems they face;
- The third is originality. Creativity is the ability to create something. The word create in this definition explains that creativity is the ability to bring up something new, for example, in the form of objects, ideas, *etc.*, which means that it is something original and has never existed before. The individual who created it may profoundly observe a problem because he sees that no one has solved it. Then, he will start looking for solutions by combining the ideas that have been there before to become new ideas (Rhoten et al., 2009; Runco, 1993).

5. Data analysis procedure

The analysis uses *Winstep* version 5.1.4 as a tool to perform analysis with the RM. First, the instrument will be analyzed for unidimensionality measures to evaluate whether the developed instrument can measure what it should measure, in this case, the student's creativity. Second, identification of the student's creativity range map by using *WrightMap* analysis is carried out so that the position of students' creativity can be known. Third, an analysis of the items was carried out by analyzing the item difficulty level, item suitability level, diagnostic rating scale, and detecting item bias. Finally, the instrument will be analyzed as a whole to determine the validity, reliability, and separation.

6. Findings

6.1. Unidimensionality

Unidimensionality analysis is carried out to determine whether the instrument can measure what it is supposed to measure. Unidimensionality analysis is done by looking at the output table 23 in *Winstep* version 5.1.4. The following Table 2 explains the unidimensionality value.

Table 2. Unidimensional analysis (source: created by authors)

Table of standardized residual variance (in eigenvalue units)				
	Empirical			Modeled
Total raw variance in observations	43.1	100.0%		100.0%
Raw variance explained by measures	23.1	53.6%		54.1%
Raw variance explained by persons	8.8	20.4%		20.6%
Raw variance explained by items	14.3	33.2%		33.5%
Raw unexplained variance (total)	20.0	46.4%	100.0%	45.9%
Unexplained variance in 1st contrast	7.7	17.9%	38.7%	
Unexplained variance in 2nd contrast	2.9	6.7%	14.4%	
Unexplained variance in 3rd contrast	1.9	4.3%	9.3%	
Unexplained variance in 4th contrast	1.5	3.4%	7.4%	
Unexplained variance in 5th contrast	1.1	2.6%	5.6%	

Based on the value of unidimensionality in Table 2, the instrument's accuracy in measuring creativity is known. It is known that the raw data variance value is 53.6%, which means that it has met the unidimensionality requirements and is included in the good category because it exceeds the 40% figure, which is the limit of the good category. In addition, the unexplained variants 1st to 5th respectively showed values of 17.9%, 6.7%, 4.3%, 3.4% and 2.6%. There is a value that exceeds the value of 15%; the unexplained 1st variant has a value of 17.9%, which means the instrument is still not perfectly measuring students' creativity.

6.2. *WrightMap* analysis (person-item map)

Based on the output of variables map table, it can be seen that the student's creativity map is in the range of –1 to 3. Item number 16 is the most challenging item for most students, and item number 1 is the most accessible item for all students. The student's average ability, which is also between logit –1 to 3, is 1.26. The average indicates that the student's ability exceeds the item difficulty level, which is at logit 0.

6.3. Item difficulty level

The difficulty level of the item can be known by looking at the item measure output table. The following is Table 3, which describes the item measure output table.

Table 3. Item measure (source: created by authors)

Entrance number	Total score	Measure	Infit		Outfit		Point measure correlation		Exact observed, %	Match expected,%
			Mean-square	*Zstandard*	Mean-square	*Zstandard*	Correlation	Expected value		
16	35	2.43	1.49	1.4	1.50	1.5	.54	.53	31.8	53.1
9	53	1.00	.54	−1.8	.53	−1.8	.73	.61	72.7	45.0
15	56	.80	.90	−.2	.84	−.5	.60	.62	50.0	43.5
20	57	.73	1.00	.1	1.00	.1	.51	.62	36.4	43.9
10	61	.48	1.42	1.4	1.45	1.5	.65	.63	36.4	44.4
8	62	.41	.61	−1.5	.64	−1.3	.68	.63	54.5	43.8
14	62	.41	1.10	.4	1.07	.3	.59	.63	50.0	43.8
13	65	.22	1.33	1.2	1.31	1.1	.50	.64	31.8	43.6
6	68	.03	.79	−.7	.81	−.6	.83	.64	36.4	42.9
12	68	.03	.97	.0	1.01	.1	.40	.64	36.4	42.9
7	71	−.16	.65	−1.3	.67	−1.3	.86	.65	36.4	42.1
11	71	−.16	1.35	1.2	1.27	1.0	.43	.65	40.9	42.1
4	74	−.35	.96	.0	.99	.0	.66	.65	27.3	41.7
19	75	−.41	1.06	.3	1.02	.2	.60	.65	45.5	41.7
5	77	−.54	.86	−.4	.86	−.4	.68	.65	59.1	42.8
18	78	−.61	1.18	.7	1.15	.6	.65	.65	27.3	42.9
3	80	−.74	.90	−.2	.90	−.3	.70	.65	50.0	43.9
2	82	−.88	1.02	.2	.95	−.1	.75	.65	49.9	45.1
17	82	−.88	1.07	.4	1.06	.3	.41	.65	49.9	45.1
1	94	−1.84	1.12	.5	.89	−.2	.79	.61	50.0	53.0
Mean	68.6	.00	1.02	.1	.99	.0			42.7	44.4
Standart deviation	12.6	.87	.25	.9	.25	.9			11.1	3.1

It is known that the standard deviation or standard deviation value is 0.87, and if it is combined with the average value, it will get a limit value for each category. Items in the very difficult category have a value above 0.87, items in the difficult category range from 0 to 0.87, items in the easy category range from 0 to −0.87, and items in the very easy category have a value less than −0.87.

Based on the division of these categories, it can be seen that the items in the creativity instrument are grouped into four categories. Items number 16 and 9 are items that fall into the very difficult category; item numbers 15, 20, 10, 8, 14, 13, 6, and 12 are items that fall into the difficult category, item numbers 7, 11, 4, 19, 5, 18, and 3 are items that fall into the easy category, and item numbers 2, 17, and 1 are items that fall into the very easy category. Thus, it can be concluded that there are two items in the very difficult category, eight items in the difficult category, seven items in the easy category, and three items in the very easy category.

6.4. Item suitability level

The level of suitability of the item will explain whether the item functions normally or is fit in taking measurements so that there is no inaccuracy in giving conclusions, in this case, is student creativity. The item fit order output table is used to determine the level of item suitability. The following is the Table 4 of fit order items.

Some requirements need to be met to find out whether an item is appropriate or not. The infit mean-square (IMS) value of each item needs to be compared with the average value plus the standard deviation value. The IMS value on each item that exceeds the average value plus the standard deviation value is said to be unsuitable. It is known that the mean value plus the standard deviation value is 1.27. If the IMS value for each item is compared with the average value plus standard deviation, it can be concluded that item numbers 16, 10, 11, and 13 are not appropriate or are declared overfit.

6.5. Rating scale diagnostic

The rating scale diagnostic is carried out to determine whether the participants understand the difference in the answer choices on the scale, in this case, the scale of 0, 1, 2, 3, 4, and 5. The rating output table (partial credit) scale (ROTPCS) determines the participants' understanding of the difference in answer choices. The following is Table 5, which describes the ROTPCS.

Detecting whether there is a difference in understanding in each answer can be seen by looking at the observed average and Andrich threshold (AT) values. If the observed average and AT values are sequentially increased, the scale of 0 to 5 can be well understood by the participants. Thus, the three alternative answers can be used.

Table 4. Item fit order (source: created by authors)

Entrance number	Total score	Measure	Infit		Outfit		Point measure correlation		Exact observed, %	Match expected, %
			Mean-square	*Zstandard*	Mean-square	*Zstandard*	Correlation	Expanded value		
16	35	2.43	1.49	1.4	1.50	1.5	.54	.53	31.8	53.1
10	61	.48	1.42	1.4	1.45	1.5	.65	.63	36.4	44.4
11	71	–.16	1.35	1.2	1.27	1.0	.43	.65	40.9	42.1
13	65	.22	1.33	1.2	1.31	1.1	.50	.64	31.8	43.6
18	78	–.61	1.18	.7	1.15	.6	.65	.65	27.3	42.9
1	94	–1.84	1.12	.5	.89	–.2	.79	.61	50.0	53.0
14	62	.41	1.10	.4	1.07	.3	.59	.63	50.0	43.8
17	82	–.88	1.07	.4	1.06	.3	.41	.65	40.9	45.1
19	75	–.41	1.06	.3	1.02	.2	.60	.65	45.5	41.7
2	82	–.88	1.02	.2	.95	–.1	.75	.65	40.9	45.1
12	68	.03	.97	.0	1.01	.1	.40	.64	36.4	42.9
20	57	.73	1.00	.1	1.00	.1	.51	.62	36.4	43.9
4	74	–.35	.96	.0	.99	.0	.66	.65	27.3	41.7
3	80	–.74	.90	–.2	.90	–.3	.70	.65	50.0	43.9
15	56	.80	.90	–.2	.84	–.5	.60	.62	50.0	43.5
5	77	–.54	.86	–.4	.86	–.4	.68	.65	59.1	42.8
6	68	.03	.79	–.7	.81	–.6	.83	.64	36.4	42.9
7	71	–.16	.65	–1.3	.67	–1.3	.86	.65	36.4	42.1
8	62	.41	.61	–1.5	.64	–1.3	.68	.63	54.5	43.8
9	53	1.00	.54	–1.8	.53	–1.8	.73	.61	72.7	45.0
Mean	68.6	.00	1.02	.1	.99	.0			42.7	44.4
Standard deviation	12.6	.87	.25	.9	.25	.9			11.1	3.1

Table 5. Rating output table (partial credit) scale (source: created by authors)

Category label	Score	Observed		Observed sample		Infit mean-square	Outfit mean-square	Andrich threshold	Category measure
		Count	%	Average	Expected				
0	0	2	0	–3.10	–1.84	.39	.57	None	(–5.35)
1	1	41	9	–.55	–.63	.94	.97	–4.23	–2.64
2	2	100	23	.42	.41	1.07	1.05	–.98	–.27
3	3	123	28	1.31	1.26	.98	.94	.65	1.14
4	4	109	25	1.82	1.96	1.27	1.22	1.74	2.43
5	5	65	15	2.77	2.65	.86	.89	2.82	(4.12)

6.6. Biased item detection

The detection of biased items determines whether there are more profitable items for individuals with specific categories than individuals with other categories. In this study, items were detected by looking at gender characteristics. The differential item functioning (DIF) output table is used to determine item bias. The following is a DIF table described in Table 6.

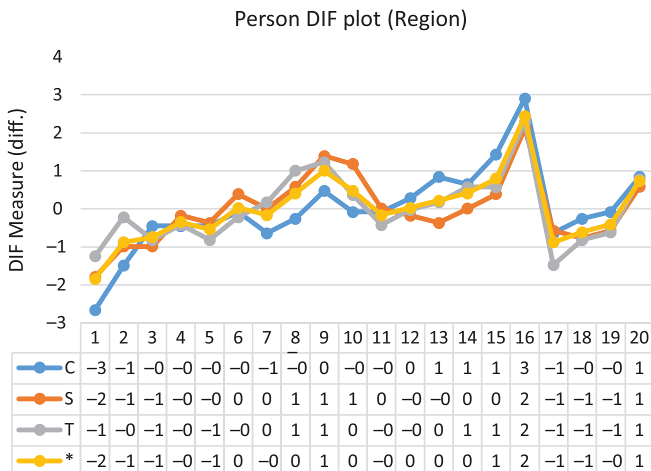
Table 6. Differential item functioning (source: created by authors)

Person classes	Summary differential item functioning			Between-class		Item number
	Chi-square	Difference	Problem	Mean-square	t = *Zstandard*	
2	.6556	1	.4181	.3431	–.1649	1
2	.6507	1	.4199	.3442	–.1633	2
2	.7498	1	.3865	.3979	–.0897	3
2	3.7344	1	.0533	2.1750	1.0986	4
2	2.7223	1	.0990	1.5351	.7972	5
2	.0564	1	.8123	.0296	–9938	6
2	.1563	1	.6926	.0819	–.7285	7
2	.1628	1	.6866	.0854	–.7156	8
2	1.8825	1	.1701	1.0398	.4992	9
2	.0973	1	.7551	.0507	–.8649	10
2	2.3233	1	.1275	1.2935	.6614	11
2	4.3235	1	.0376	2.5679	1.2550	12
2	1.1268	1	.1730	1.0210	.4861	13
2	1.8569	1	.2885	.6065	.1457	14
2	3.4565	1	.0630	2.0009	1.0232	15
2	.9215	1	.3371	.4967	.300	16
2	.0514	1	.8206	.0261	–1.0205	17
2	.8528	1	.3558	.4542	–.0193	18
2	.5874	1	.4434	.3108	–.2130	19
2	.6108	1	.4345	.3248	–.1918	20

An item is said to be biased if the probability value of the item, as shown in Table 6, is below 5% or 0.05. Table 6 analyzes the bias in terms of gender category. It is known that Item number 12 has a value of 0.0376 which means the value is below 5% or 0.05. Therefore, item number 12 can be said to be biased when viewed from the gender category.

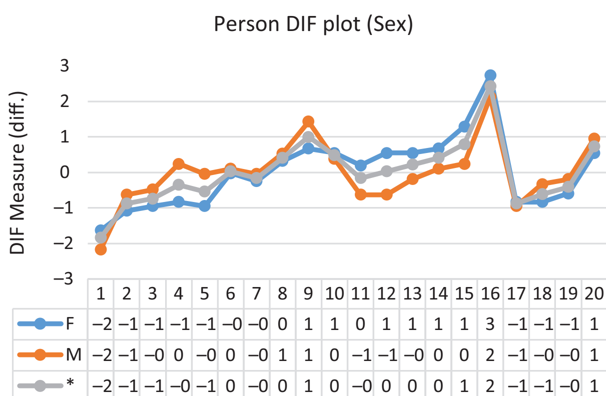
A graphic image of DIF regarding the logit position for each item by region and gender can be seen in Figure 1.

From the Figure 1, it can be concluded that there is no bias in the regional category, but in the gender category, there is a bias in item number 12. This item is easier for male participants to do. This can be seen from the male DIF score, whose position is far below the female category, which means that male participants are much easier to work on question number 12 than female participants.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
C	−3	−1	−0	−0	−0	−0	−1	−0	0	−0	−0	0	1	1	1	3	−1	−0	−0	1
S	−2	−1	−1	−0	−0	0	0	1	1	1	0	−0	−0	0	0	2	−1	−1	−1	1
T	−1	−0	−1	−0	−1	−0	0	1	1	0	−0	−0	0	1	1	2	−1	−1	−1	1
*	−2	−1	−1	−0	−1	0	−0	0	1	0	−0	0	0	0	1	2	−1	−1	−0	1

Note: C = Cirebon, West Java; S = Sukabumi, West Java; T = Tasikmalaya, West Java.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
F	−2	−1	−1	−1	−1	−0	−0	0	1	1	0	1	1	1	1	3	−1	−1	−1	1
M	−2	−1	−0	0	−0	0	−0	1	1	0	−1	−1	−0	0	0	2	−1	−0	−0	1
*	−2	−1	−1	−0	−1	0	−0	0	1	0	−0	0	0	0	1	2	−1	−1	−0	1

Note: F = female; M = male.

Figure 1. Differential item functioning graphic (source: created by authors)

6.7. Instrument analysis

Instrument analysis is carried out by paying attention to the output table of the statistical summary. The following is a summary of the statistics described in Table 7.

Table 7. Summary statistic (source: created by authors)

Summary person								
	Total			Model	Infit		Outfit	
	Score	Count	Measure	Standard error	Mean-square	*Zstandard*	Mean-square	*Zstandard*
Mean	62.3	20.0	1.26	.27	1.00	–.1	.99	–.1
Standard deviation	13.5	.0	.99	.02	.47	1.5	.48	1.5
Maximum	84.0	20.0	2.89	.32	2.25	3.2	2.38	3.5
Minimum	32.0	20.0	–1.13	.26	.37	–2.8	.38	-2.7
Real root-mean-square deviation	.30	True standard deviation	.95	Separation	3.16	Person reliability		.91
Model root-mean-square deviation	.27	True standard deviation	.96	Separation	3.48	Person reliability		.92
Standard error of person mean = .22								

Person raw score-to-measure correlation = 1.00

Cronbach's alpha = 20 person raw score ("test") reliability = .92

Summary item								
	Total			Model	Infit		Outfit	
	Score	Count	Measure	Standard error	Mean-square	*Zstandard*	Mean-squre	*Zstandard*
Mean	68.6	22.0	.00	.26	1.02	.1	.99	.0
Standard deviation	12.6	.0	.87	.02	.25	.9	.25	.9
Maximum	94.0	22.0	2.43	.31	1.49	1.4	1.50	1.5
Minimum	35.0	22.0	–1.84	.25	.54	–1.8	.53	–1.8
Real root-mean-square deviation	.28	True standard deviation	.83	Separation	2.99	Item reliability		.90
Model root-mean-square deviation	.26	True standard deviation	.83	Separation	3.18	Item reliability		.91
Standard error of person mean = .20								

The items' difficulty level on the participants can be seen in the average value of the person measure compared to the average value of the item measure. It is known that the average value of the person measure is 1.26 logit, and the average value of the item measure is 0.0 logit. This shows that the participants' ability is higher than the difficulty of the items on the instrument.

There is reliability that can be seen in Table 7, ranging from instrument reliability, person reliability, and item reliability. The Cronbach's alpha value indicates the instrument's reliability, that is, the interaction between the person and the item as a whole. In Table 7, it is known that the Cronbach's alpha value is 0.92, which means that the instrument's reliability is in the very good category because it exceeds the very good category limit, which is 0.8. The person reliability value shows a value of 0.91, which means it is included in the very good category because the value is in the range of 0.91 to 0.94. The value of item reliability shows a value of 0.90 which means item reliability is included in the good category because the value is in the range of 0.81 to 0.90.

Other data that can be used are the IMS and outfit mean-square (OMS) values. In the person table, it is known that the IMS and OMS values are 1.00 and 0.99, respectively, and the items table shows that the IMS and OMS values are 1.02 and 0.99, respectively. The ideal value for IMS and OMS in the person and item tables is 1, the closer to 1, the better. Thus, it can be concluded that the IMS and OMS values in the person and item tables can be said to be good because they are close to 1.

Next, the infit *Zstandard* and outfit *Zstandard* data are in the person table and the item table. The infit *Zstandard* and outfit *Zstandard* values in the person table are –0.1 and –0.1, respectively, and the infit *Zstandard* and outfit *Zstandard* values in the item table are 0.1 and 0. The ideal values for infit *Zstandard* and outfit *Zstandard* in the person and item tables are 0, the closer to 0, the better. Thus, it can be concluded that the infit *Zstandard* and outfit *Zstandard* values in the person and item tables can be said to be good because they are close to 0.

In the person table and the item table, there is a separation value. These values can identify groups of respondents and items. The equation used to see the grouping more closely is called the strata separator. The strata separator equation is $H = \{(4 \times \text{separation}) + 1\} / 3$. The greater the strata separator value, the better the quality of the instrument in terms of overall respondents and items because it can determine the number of respondents. It is known that the separation value in the person table is 3 (after rounding up), and the separation value in the item table also has the same value, 3 (after rounding). Therefore, after the separation value in the person and item tables is entered into the strata separator equation, the separation value for persons and items is 4.3, and if it is rounded up to 4. Then, it is known that the respondent groups are categorized into four groups.

Discussion

This study aims to analyze the validity of ICS by using the IRT model: the RM. The analysis results show that several important findings can provide an evaluation note of the instrument. The dimensionality of the ICS, especially in the unexplained 1st variant, shows that the ICS still does not fully measure creativity. In the analysis of the items suitability level that identifies the function of the item, whether it is functioning normally or not, it is found that item numbers 10, 11, 13, and 16 are declared to overfit or function abnormally. In addition, some items are biased in the gender category, that is, item number 12. Thus, the items will give unbalanced results because the items will tend to be easier for men than women. Even so, the instrument's reliability is in the very good category, and the reliability of the item is in a good category, meaning that if the ICS is used repeatedly in the same group, it will not produce much significant difference in information.

The ICS unidimensionality test showed that the ICS still did not measure what it was supposed to measure. Dimensionality is a fundamental measure to determine the construct validity of an instrument (Duruturk et al., 2015; Linacre, 1998). The items' dimensionality on the ICS should have a latent dimension that measures what should be measured (Ul Hassan & Miller, 2022). However, it should be understood that in the measurement process, each individual uses his respective abilities to understand how an item should be answered (Muslihin et al., 2022; Linacre, 1998; Tennant et al., 2004). The ICS dimensionality results show that the ICS instrument still measures multidimensionally, meaning that the items used in the instrument still describe other dimensions outside the creativity dimension.

Some items on the ICS have mismatched or do not function normally. In the ICS instrument, the misfit items are the overfit items, meaning that these items cannot provide information about how different each individual's ability is (Andrich, 2010; Fisher, 2007; Stout et al., 2012). Overfit items may be easy to do because the answers to these items already exist in other items (Linacre, 2000). Improvements can be made by eliminating items (if other items represent indicators in the creativity variable) or improving the sentences used for these items.

The item bias on the ICS instrument indicates that the item benefits individuals with certain categories. Biased items will discriminate against individuals with other categories to have different results (Mellenbergh, 1989). The individual's ability level will be non-objective because biased items have advantages for specific categories. Unfortunately, in this study, item bias analysis was only conducted based on gender and region categories. Various categories can be used, such as age, race, academic achievement index, activity, *etc.*, to identify more in-depth item bias so that the measurement will get more objective results.

Based on these findings, it can be seen that the instrument requires evaluation to obtain a more accurate measuring instrument. However, it should also be understood that the measure of validity carried out with the RM is a measure of construct validity, meaning that the measure is only based on the theoretical context (Pallant & Tennant, 2007; Smiley, 2015). Concurrent validity and predictive validity may need to be done. That way, the instrument's validity will be clarified by comparing the results given from the instrument with the conditions given by other measuring instruments with the same concept (Gasser et al., 2007; Reddy et al., 2013). Therefore, after evaluating the instrument, it is hoped that research will be carried out to determine how the instrument's validity is seen from the concurrent validity and predictive validity.

Conclusions

The results of the construct analysis provide some notes for evaluating the ICS instrument. The results given by the ICS measuring instrument are still not perfectly measuring creativity. ICS still measures multidimensionally, or in other words, there are still ICS' dimensions that do not measure creativity. In testing the suitability of items on the ICS, there are still items said to overfit. There may be other items that indicate the answers to the four numbers so that the items are said to overfit. Items that are said to be biased also still exist when viewed from the gender category. Thus, the group will be measured evenly and not discriminate against each other. It is hoped that the measurement of bias items in future research can provide more explanation by comparing it with other categories such as age, race, academic achievement index, activity, *etc.*, as an instrument evaluation material to provide more objective results.

Nevertheless, the instrument has reliability that can be said to be good. The instrument's reliability is in the very good category, and the reliability of the item is in a good category. Hence, based on the analysis results, it is hoped that an evaluation of the instrument will be carried out, and it can be retested to get more objective results. Validity testing can also be done with other validity tests, such as concurrent validity and predictive validity.

Further research will be tested on more diverse subjects. Indonesia with various ethnic cultures is an opportunity to develop instruments. This test will provide more information about the quality of the instrument's measurement of different populations. That way the instrument can be evaluated to get a measuring instrument that measures accurately.

References

Andrich, D. (2010). Rasch models. In P. Peterson, E. Baker, & B. McGaw (Eds.-in-Chief), *International encyclopedia of education* (pp. 111–112). Elsevier Science. https://doi.org/10.1016/B978-0-08-044894-7.00258-X

Andrich, D. (1988). *Quantitative applications in the socials sciences. Rasch models for measurement. A Sage University Paper No. 68*. Sage Publications, Inc. https://doi.org/10.4135/9781412985598

Bond, T. G., & Fox, Ch. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

Boone, W. J., Staver, J. R., & Yale, R. S. (2014). *Rasch analysis in the human sciences*. Springer Science+Business Media. https://doi.org/10.1007/978-94-007-6857-4

Braslauskas, J. (2021). Developing intercultural competences and creativity: The foundation for successful intercultural communication. *Creativity Studies, 14*(1), 197–217. https://doi.org/10.3846/cs.2021.14583

Chakrabarti, A., & Bligh, Th. P. (1994). An approach to functional synthesis of solutions in mechanical conceptual design. Part I: Introduction and knowledge representation. *Research in Engineering Design, 6*, 127–141. https://doi.org/10.1007/BF01607275

Chiu, F.-Ch., Hsu, Ch.-Ch., Lin, Y.-N., Liu, Ch.-H., Chen, H.-Ch., & Lin, Ch.-H. (2019). Effects of creative thinking and its personality determinants on negative emotion regulation. *Psychological Reports, 122*(3), 916–943. https://doi.org/10.1177/0033294118775973

Departemen Pendidikan Nasional. (2008). *Penataan Pendidikan Profesional Konselor dan Layanan Bimbingan dan Konseling dalam Jalur Pendidikan Formal*. Departemen Pendidikan Nasional.

Dreu, de C. K. W., Nijstad, B. A., & Baas, M. (2011). Behavioral activation links to creativity because of increased cognitive flexibility. *Social Psychological and Personality Science, 2*(1), 72–80. https://doi.org/10.1177/1948550610381789

Du, Y., Xie, L., An Zhong, J., Zou, H., Law, R., & Yan, X. (2019). Creativity fostering teacher behavior on student creative achievement: Mediation of intrinsic motivation and moderation of openness to experience. *School Psychology International, 40*(5), 525–542. https://doi.org/10.1177/0143034319868271

Duruturk, N., Tonga, E., Gabel, Ch. Ph., Acar, M., & Tekindal, A. (2015). Cross-cultural adaptation, reliability and validity of the Turkish version of the lower limb functional index. *Disability and Rehabilitation, 37*(26), 2439–2444. https://doi.org/10.3109/09638288.2015.1024342

Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (Eds.). (2020). *Global Innovation Index 2020: Who will finance innovation?* Cornell University, INSEAD, and the World Intellectual Property Organization.

Engelhard, Jr. G., & Wind, S. A. (2018). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Routledge. https://doi.org/10.4324/9781315766829

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

Fuchs-Beauchamp, K. D., Karnes, M. B., & Johnson, L. J. (1993). Creativity and intelligence in preschoolers. *Gifted Child Quarterly, 37*(3), 113–117. https://doi.org/10.1177/001698629303700303

Gasser, C. E., Larson, L. M., & Borgen, F. H. (2007). Concurrent validity of the 2005 strong interest inventory: An examination of gender and major field of study. *Journal of Career Assessment, 15*(1), 23–43. https://doi.org/10.1177/1069072706294516

Glăveanu, V. P. (2010). Paradigms in the study of creativity: Introducing the perspective of cultural psychology. *New Ideas in Psychology, 28*(1), 79–93. https://doi.org/10.1016/j.newideapsych.2009.07.007

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444–454. https://doi.org/10.1037/h0063487

Guilford, J. P. (1962). Potentiality for creativity. *Gifted Child Quarterly, 6*(3), 87–90. https://doi.org/10.1177/001698626200600307

Guilford, J. P. (1959). Traits of creativity. In H. H. Anderson (Ed.), *Creativity and its cultivation* (pp. 142–161). Harper & Row.

Higgins, G. E. (2007). Examining the original Grasmick scale: A Rasch model approach. *Criminal Justice and Behavior, 34*(2), 157–178. https://doi.org/10.1177/0093854806290071

Kowsar, K. S. Sh., & Mukherjee, S. (2021). Recreating *Hamlet*: Creativity of Vishal Bhardwaj in *Haider*. *Creativity Studies, 14*(1), 90–98. https://doi.org/10.3846/cs.2021.11556

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement, 2*(3), 266–283.

Linacre, J. M. (2000). *Fit diagnosis: Infit outfit mean-square standardized*. https://www.winsteps.com/winman/misfitdiagnosis.htm

Long, B. H., & Henderson, E. H. (1965). Opinion formation and creativity in elementary school children. *Psychological Reports, 17*(1), 219–223. https://doi.org/10.2466/pr0.1965.17.1.219

Mellenbergh, G. J. (1989). Item bias and Item response theory. *International Journal of Educational Research, 13*(2), 127–143. https://doi.org/10.1016/0883-0355(89)90002-5

Mohamad, M. M., Sulaiman, N. L., Sern, L. Ch., & Salleh, K. M. (2015). Measuring the validity and reliability of research instruments. *Procedia – Social and Behavioral Sciences, 204*, 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

Muslihin, H. Y., Suryana, D., Ahman, A., Suherman, U., & Hayati Dahlan, T. (2022). Analysis of the reliability and validity of the self-determination questionnaire using Rasch Model. *International Journal of Instruction, 15*(2), 207–222. https://doi.org/10.29333/iji.2022.15212a

Nurhudaya, N., Taufik, A., Sakti Yudha, E., & Suryana, D. (2019). The Raven's advanced progressive matrices in education assessment with a Rasch Analysis. *Universal Journal of Educational Research, 7*(9), 1996–2002. https://doi.org/10.13189/ujer.2019.070921

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the hospital anxiety and depression scale (HADS). *British Journal of Clinical Psychology, 46*(1), 1–18. https://doi.org/10.1348/014466506X96931

Quiñones-Gómez, J. C. (2021). Creativity forward: A framework that integrates data analysis techniques to foster creativity within the creative process in user experience contexts. *Creativity Studies, 14*(1), 51–73. https://doi.org/10.3846/cs.2021.12933

Reddy, L. A., Fabiano, G. A., & Dudek, Ch. M. (2013). Concurrent validity of the classroom strategies scale for elementary school – observer form. *Journal of Psychoeducational Assessment, 31*(3), 258–270. https://doi.org/10.1177/0734282912462829

Rhoten, D., O'Connor, E., & Hackett, E. J. (2009). The act of collaborative creation and the art of integrative creativity: Originality, disciplinarity and interdisciplinarity. *Thesis Eleven, 96*(1), 83–108. https://doi.org/10.1177/0725513608099121

Runco, M. A. (1993). Operant theories of insight, originality, and creativity. *American Behavioral Scientist, 37*(1), 54–67. https://doi.org/10.1177/0002764293037001006

Sen, A. K., & Hagtvet, K. A. (1993). Correlations among creativity, intelligence, personality, and academic achievement. *Perceptual and Motor Skills, 77*(2), 497–498. https://doi.org/10.2466/pms.1993.77.2.497

Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies, 24*(2), 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0

Sisk, D. A. (2014). Optimizing creativity. *Gifted Education International, 30*(2), 148–159. https://doi.org/10.1177/0261429413481124

Smiley, J. (2015). Classical test theory or Rasch: A personal account from a novice user. *Shiken, 19*(1), 16–29.

Sternberg, R. J. (2003). A broad view of intelligence: The theory of successful intelligence. *Consulting Psychology Journal: Practice and Research, 55*(3), 139–154. https://doi.org/10.1037/1061-4087.55.3.139

Stout, J. L., Gorton, III G. E., Novacheck, T. F., Bagley, A. M., Tervo, R. C., Bevans, K., & Tucker, C. A. (2012). Rasch analysis of items from two self-report measures of motor function: Determination of item difficulty and relationships with children's ability levels. *Developmental Medicine and Child Neurology, 54*(5), 443–450. https://doi.org/10.1111/j.1469-8749.2012.04231.x

Tennant, A., McKenna, S. P., & Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health, 7*(S1), 22–26. https://doi.org/10.1111/j.1524-4733.2004.7s106.x

Ul Hassan, M., & Miller, F. (2022). Discrimination with unidimensional and multidimensional Item response theory models for educational data. *Communications in Statistics: Simulation and Computation, 51*(6), 2992–3012. https://doi.org/10.1080/03610918.2019.1705344

Wilson, H. E. (2007). Creative problem solving kit. *Journal for the Education of the Gifted, 31*(1), 110–116. https://doi.org/10.4219/jeg-2007-515

Yamaoka, A., & Yukawa, Sh. (2020). Does mind wandering during the thought incubation period improve creativity and worsen mood? *Psychological Reports, 123*(5), 1785–1800. https://doi.org/10.1177/0033294119896039

Yasin, R. M., Nur Yunus, F. A., Che Rus, R., Ahmad, A., & Bekri Rahim, M. (2015). Validity and reliability learning transfer item using Rasch measurement model. *Procedia – Social and Behavioral Sciences, 204*, 212–217. https://doi.org/10.1016/j.sbspro.2015.08.143

Zosh, J. M., Hirsh-Pasek, K., Michnick Golinkoff, R., & Dore, R. A. (2017). Where learning meets creativity: The promise of guided play. In R. A. Beghetto & B. Sriraman (Eds.), *Creativity theory and action in education. Vol. 1. Creative contradictions in education: Cross disciplinary paradoxes and perspectives* (pp. 165–180). Springer. https://doi.org/10.1007/978-3-319-21924-0_10