DEVELOPING A SCIENTIFIC CREATIVITY TEST FOR FIFTH GRADERS

Nyet Moi Siew, Chin Lu Chong, Kim On Chin
Sabah University of Malaysia, Malaysia
E-mail: snyetmoi@yahoo.com, ev1lyn87@hotmail.com, kimonchin@gmail.com

Abstract

Testing creativity in general has been well researched, but little has been reported on the development of instruments to test scientific creativity among primary school students. This study describes the development and validation of scientific creativity test for primary school fifth graders. A Scientific Creativity Structure Model (SCSM) was used to guide the development of test items through three dimensions called trait, process and product. Torrance Tests of Creative Thinking (TCTT) were used to evaluate test item answers. Two equivalent, parallel scientific creativity tests were developed each of which consisted of 4 items posed in the form of: technical product, advances in science or scientific knowledge, understanding of scientific phenomenon and scientific problem solving. The scientific creativity test was validated through analysis of item response data of 206 fifth grade students from two Malaysian primary schools. The scientific creativity test was found to have high internal consistency, inter-scorer reliability and face validity. Both Form A and Form B of the test had an acceptable discrimination index range. The test showed a weak positive, but significant correlation between the items in product and process dimensions, but a very strong correlation between the three trait dimensions of SCSM. Test items on science problem solving have a strong indication loading on spatial analytical thinking. The Item analysis suggests that this test would be useful in assessing scientific creativity of the fifth grade students with further review on test items measuring science problem solving.
Key words: *item analysis, primary school students, scientific creativity test, Scientific Creativity Structure Model, Torrance Tests of Creative Thinking.*

Introduction

Infusing creativity elements into science, technology and engineering classrooms is an important movement in educational reforms in most countries. According to Özbey & Alisinanoğlu (2008), science education is necessary in the primary school period for children to improve their creativity and learn about different perspectives. Children aged 7–12 need science education to learn about their environment and natural occurrences, and to generate original ideas (AAAS, 2003). The children are expected to develop scientific ideas that are original, useful, and surprising as what Simonton (2012) defined as scientific creativity.

There is little evidence that instructions are being used to assess children' scientific creativity. Kozma reports, that "there are few studies that go beyond traditional measures of student learning to include outcomes such as creativity, complex problem solving, collaboration, and the ability to learn, and even fewer that do this in the context of developing countries" (2010: 17). Much of the difficulty lies in the lack of a test to assess children's scientific creativity in primary school science classrooms. Torrance (1974) has developed a test of creative thinking (TCTT), which is widely used to measure the creativity traits of an individual in terms of fluency, flexibility, originality, and elaboration. However, TCTT is mainly used to measure the general creativity of an individual (Kim, 2006). Therefore, a scientific creativity test is required to measure how children's scientific ideas are developed in original, useful and surprising ways.

The present study addressed this concern by developing a scientific creativity test for primary school teachers to measure scientific creativity among fifth grade students in Malaysia.

Scientific Creativity

Hu & Adey (2002) elaborated a set of hypotheses about the structure of scientific creativity. They infer that scientific creativity is different from general creativity since it is concerned with creative science experiments, creative scientific problem finding and solving, and creative science activity. They conclude that scientific creativity can be defined as the ability to use scientific knowledge and skills to produce a certain product that is original and has certain social or personal value. Thus, the concept of 'scientific' or 'science' and 'creativity' are not segregated but fully integrated as a whole new term that is an accurate representative of the idea about students' creativity in science and with science.

Engineering and pharmaceutical drugs (conventional medicines) have contributed to the further elaboration on the practical definition of scientific creativity. Engineering, by definition, is the application of scientific and relevant practical knowledge to invent, design, improve and maintain machinery, devices, systems, materials and structure (Engineers' Council for Professional Development, 1947). Indeed, the field of engineering is founded based on the idea of 'scientific creativity,' where engineering uses scientific knowledge to create something original with the given value, such as a pharmaceutical drug. A lot of technological advancements are based on scientific creativity of individuals or groups of people in their specialised field. If a science teacher considers the child's world from this perspective, the ideas within the term 'scientific creativity' can be seen with greater clarity as physical, real-world examples, rather than abstract descriptions.

However, primary school students are not equipped with advanced scientific knowledge. This means that 'scientific creativity' at their level should be based on accessible knowledge, such as basic concepts of gravity, photosynthesis, or simple measurement and scientific investigation. A key precept of scientific creativity in primary school students is the ability of students to use basic scientific knowledge to produce simple, original products (not necessarily physical) that meet certain value(s).

Scientific Creativity Test

Research has been done on instruments or tests developed to test scientific creativity among school students. For example, Aktamıs, et al. (2005) developed a test to measure scientific creativity and scientific process skills for primary school students in Turkey. The researchers used Scientific Creativity Structure Model (SCSM) proposed by Hu & Adey (2002), with minor changes to suit Turkish language and culture. SCSM (Figure 1) is a theoretical model consisting of three dimensions, the process dimension (thinking and imagination), the trait dimension (fluency, flexibility, and originality), and the product dimension (technical product, science knowledge, science phenomena, and science problem). Their pilot test was found to have adequate reliability and validity. Unfortunately, they do not report the testing of their instrument with primary school students, so, the utility of this instrument cannot yet be determined.

Other research by Sahin-Pekmez, Aktamis, & Can (2009), adapted Hu & Adey's (2002) SCSM into Turkish for reliability analysis. After using the test on 7th Grade Students in selected schools, most students' answers were about how to achieve the goal asked in each test item, rather than the mechanism on how to achieve the goal itself. For instance, for the item which asked the students to design an apple-picking machine; point out the name and function of each part. Most students gave the answers like "the machine will find apple, pick it, transport it to the basket". Only a handful tried to elaborate scientifically how the suggested new design would make the machine work. This suggests that such item does not address clearly the important aspect that needs to be evaluated in that item or it could be caused by a certain unknown mental tendency of most students to focus on the goal instead of the scientific creativity to achieve such goal. Either way, the former suggestion could be something that needs to be addressed in this research.

Rawat (2010) used Sharma and Shukla's (1986) Verbal Test of Scientific Creativity and three-way ANOVA to study the fluency component of creative scientific talent of 1120 elementary students in Himachal Pradesh. The results showed significant differences between urban and rural areas with urban areas having better fluency than the rural area. The strength of this research is its high sample spaces in different areas, which translate to more accurate generalized findings. Unfortunately, Rawat (2010)'s study focused specifically on the fluency component of scientific creativity. Therefore, an instrument that tests various components of scientific creative talent needs to be applied to primary school students.

Rosen & Mosharraf (2014) use Online Performance Assessment to explore patterns in performance and motivation in creativity computer-based prototype assessment among 14 year-old students from the United States, United Kingdom, Turkey, and South Africa. While the assessment is developed from previous research of the same kind, their research focuses more on creativity in general, not on scientific creativity specifically. Also, due to a rather small sample of 87 students from four countries, their findings cannot be generalized and extrapolated.

To sum up, useful research has been conducted on instruments or tests for creativity, but only a handful of instruments have tested scientific creativity in primary schools. It is relatively difficult to test scientific creativity compared to creativity in general, since scientific creativity requires the uniqueness and appropriateness of the student's responses to questions related to science (Mohamed, 2006).

Research findings showed that students with low levels of scientific knowledge tend to answer the items without using science, rendering the data useless, there are even less potential samples compared to testing out creativity in general (Sahir-Pekmez, Aktamis & Can, 2009). However, if scientific creativity is to be an important element in primary education as stated in Malaysia Education Blueprint 2013-2025, then it is important to have a reliable instrument that can measure scientific creativity (formative and summative) among primary school students. Thus, this study describes such a test for scientific creativity among the fifth grade students. According to Piaget's developmental theory, the formal operations stage starts from the age of 11 or 12 to adulthood. The fifth graders at the age of 11 are likely to make the transition from concrete operation stage to formal operational thinking. During the transition stage, children develop the ability to think in a logical way (Inhleder & Piaget, 1958). The ability to formulate hypotheses is one of the most important processes of logical thought (Wolfinger, 2000). Aktamıs, et al. (2005) considered the ability to formulate hypotheses as an essential component of scientific creativity. At this stage, the fifth graders acquire the ability to formulate hypotheses and solve problems through producing several possible methods when they do science. Thus, assessing the scientific creativity of the fifth graders is worth attention because it can help science educators provide the appropriate teaching methods that promote scientific creativity.

The Development of Scientific Creativity Test

A scientific creativity test, inspired by the SCSM proposed by Hu & Adey (2002), was developed for the fifth grade students. SCSM was based on Guilford's (1959) Structure-of-Intellect model. The SCSM (Figure 1) was designed as a theoretical foundation to measure scientific creativity through three dimensions called trait, process and product. The three dimensional model of scientific creativity offers 24 cells (4 product dimension x 3 trait dimension x 2 process dimension) for each of items may be designed. Validity of SCSM was established through several studies (Aktamıs, et al., 2005; Pekmez, Aktamis, & Taskin, 2009).

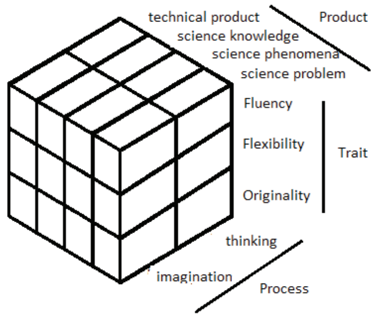

Figure 1: The Scientific Structure Creativity Model (SSCM) (Hu & Adey, 2002).

The trait dimension in SCSM is said to reflect creative personality traits inherent in all individuals. Its three central features are fluency, flexibility and originality. Fluency is defined as the number of ideas produced. In the present research context, the more scientific ideas that are produced by a student, the greater fluency that student has. Flexibility is defined as the ability to not be tied to orthodox approaches if they are no longer efficient. This means a student with greater flexibility traits tend to search for ideas with different categories or approaches. Lastly, originality is defined as an answer that is statistically rare, occurring only occasionally in a given population sample. Therefore, a student that gives rare, but appropriate answers to a question is considered to have higher originality traits than the other students.

The process dimension of SCSM reflects a series of intellectual mental operations by an individual to produce creative product(s) with their trait dimension. Its two central features are called creative imagination and creative thinking. Creative imagination is associated with using explorative mental operations that lead to new and related ideas (Craft, 2000; Sefertzi, 2000; Smith, Gerkens, Shah, Vargas-Hernandez, 2003). Creative thinking involves divergent thinking defined by Guilford (1967) as thinking in various directions in order to arrive at alternative solutions to a problem. In this study, students were required to generate as many novel outcomes as he or she could to a certain prompt.

The product dimension of SCSM is in the form of a technical product, advances in scientific knowledge, understanding of scientific phenomenon, or scientific problem solving. A technical product refers to a science-based tool that is technologically engineered to perform specific tasks and is subject to innovation. Scientific knowledge refers to knowledge in any science-based field such as Physics, Biology, Chemistry, Geology, Engineering, and others. Scientific knowledge, by definition, is the knowledge gained by systematic study through scientific methods, based on observable and measurable evidence (Wilson, 1998) and accepted by the scientific community. The domain-relevant skills of Amabile's (1983) model of creativity were used as a major theoretical basis in measuring scientific knowledge which represented a good amount of scientific knowledge that an individual should know in order to be considered creative in a specific domain.

Scientific phenomena, by definition, are natural physical phenomena that can be explained scientifically. This refers only to observable and measurable phenomena (e.g. hurricanes, whirlpools, earthquakes), not to paranormal phenomena (e.g. ghosts and UFOs). According to Johnston (2005), the scientific concept the students develop everyday are relevant to the scientific phenomena that they experienced in their world.

Science problems refer to issues that require scientific knowledge to be solved. The problems do not necessarily require high level scientific knowledge. By taking into account the level of children cognitive development, basic scientific knowledge based on real-life experience will suffice, for example; we fall back down after we jump (gravity), we walk from here to there (measurement), and we question or test different ways to explore a problem (scientific inquiry). According to Lubart (1994), if there is a problem, then there is the possibility of creative solutions. Thus, by offering a scientific problem to students, there is a possibility for them to produce a creative scientific solution.

The Usage of SCSM and Torrance's Test of Creative Thinking (TTCT) in Developing Test Items

The basis for evaluation of TTCT is similar to the trait dimension of SCSM with the addition of an Elaboration component. However, in order for TCTT to accommodate the three dimensions of SCSM, the Elaboration component of TTCT is not evaluated in the development of test items. Therefore, excluding the Elaboration component of TCTT leaves just the three dimensions of SSCM. With this in mind, the researchers selected SCSM as a model to guide the development of test items for the fifth grade students, while using TCTT as a guide to evaluate the creativity indexes of fluency, flexibility and originality in student's answers in test items.

Pilot Test Results

A total of 24 cells are available in SCSM, providing 24 different aspects of scientific creativity that can be tested. The researchers tested all 24 cells, with one item for each cell which were revised by science education researchers and primary school science teachers in Sabah. The results showed that only nine cells were considered relevant to primary school student's level of scientific knowledge. However, as the pilot test was on the fifth grade students, the results showed that only 4 items were suitable for their level of scientific knowledge and cognitive development, with regard to age, language proficiency and relevance to their living environment. Therefore, the researchers designed two equivalent, parallel scientific creativity tests each of which consisted of 4 items posed in the form of:

- Item 1: technical product.
- Item 2: advances in science or scientific knowledge.
- Item 3: understanding of scientific phenomenon.
- Item 4: scientific problem solving.

By employing two equivalent scientific creativity tests**,** the test items that gave the most consistent results were proposed to teachers to test students' scientific creativity.

Technical Product (Item 1)

During the development of test item 1 for each parallel test, the researcher used Torrance's Product Improvement Task as a guide to create an item that measured the student's fluency, flexibility, and originality through the student's creative imagination and thinking to develop a technical product. This tested a total of six SCSM cells (3 trait dimension x 2 process dimension x 1 product dimension). The Form A and Form B test items 1 were designed as follows:

> *'Suggest as many scientific improvements to a pen (Form A) /whiteboard pen (Form B) to make it look interesting, unusual and no need to be practical. You can show your idea using a drawing.'*

In the present research, the technical product chosen is a pen (in Form A) and a marker pen (in Form B). A pen and marker pen is a science-based tool that is technologically engineered to perform specific tasks and is subject to innovation. Therefore, the students would show their creativity to improve the given technical product in both primary and auxiliary purposes by illustrating it through a drawing with its labelled compartment.

To measure the ability of students to improve a technical product that is designed using scientific principle, the answers provided by the students based on the question given were analysed and compartmentalized into a different category. Students who gave a wide variation in their design would have better fluency, flexibility and originality in their creative thinking.

Science Knowledge (Item 2)

The test item 2 measured the student's fluency, flexibility and originality through practising the student's creative thinking to demonstrate their scientific knowledge. To demonstrate the fifth grader's science knowledge, a brief description of scientific terms associated with the year five science syllabus was requested. Hence, this test item tested three cell of SCSM (3 trait dimension x 1 process dimension x 1 product dimension). Form A and Form B items 2 were designed as follows:

'Write down as many scientific words as you know about 'magnet' (Form A) /'microorganisms' (Form B)

To sum up, students' creative thinking was measured and evaluated by how original and related the scientific terms were to the term 'magnet' and 'microorganism'. A student needed to generate as many responses as he or she could on four categories: uses, characteristics, types, and effects.

Science Phenomena (Item 3)

To measure students' imagination in explaining a scientific phenomenon, students were asked to write a story to explain the possible implication of the phenomena "sun loses its light" and "plants can move like animals," based on their connection and experience with natural phenomena. The students' scientific imagination was evaluated by the quantity and relevance of their story to phenomena.

Test item 3 measured students' fluency, flexibility and originality through their imagination in demonstrating understanding of scientific phenomena. It tested three SCSM cells (3 trait dimension x 1 process dimension x 1 product dimension). Form A and Form B items 3 were:

Write as much as possible in an interesting scientific story to imagine the following topics:-
'The sun is losing its light' (Form A)
'Plants can move like animals' (Form B)

Science Problem (Item 4)

The test item 4 measures students' flexibility and originality through creative thinking and imagination to solve scientific problems. It tested four SCSM cells (2 trait dimension x 2 process dimension x 1 product dimension). Form A and Form B items 4 were:

By using as many methods as possible, divide a square into 4 equal parts (same form). Show your answer using a drawing' (Form A).

'By rearranging or removing matchsticks of the following symbols, create as many symbols as possible by using 5 matchsticks' (Form B).

The quantity of symbols created and different methods to divide a square shown by the students would determine the creative science problem solving ability of the students.

Scoring Procedures

Item 1 to 3 scores were the sums of fluency, flexibility, and originality scores. Table 1 shows the scoring criteria used for assessing the creativity indexes. The fluency score was obtained by counting all the separate student responses regardless of quality. The flexibility score for each task was obtained by counting the number of categories given in the answer. Frequencies and percentages of each response were computed to obtain the originality score. If the probability of a response was smaller than 5%, 2 points were given; for 5-10% probability, 1 point was given; for > 10%, 0 points were given.

The item 4 score was the sum of the flexibility and the originality scores. The flexibility score for item 4 was obtained by counting the number of methods or created symbols given in the answer. As there was only one method or symbol in each response, flexibility represented the same scoring of fluency, thus only flexibility was included in the scoring. The task 4 score was computed again by tabulating all answers of all students, and then rating a particular answer for its originality score as employed by Hu and Adey (2002). A probability < 5% received 3 points; probabilities of 5 -10 got 2 points; probability > 10 got 1 point. There was one score for each method of division (Form A) and creating symbols (Form B) in task 4. Most students got 2 or 3 points, some got 15 points. Generally, it was impossible for a student to get 0 points because there were at least 2 or 3 very simple divisions in Form A and symbols creation in Form B.

Table 1. Scoring criteria for creativity indexes (Adapted and adopted from Torrance, 1990; Hu and Adey, 2002).

Creativity indexes	Scoring criteria	Score awarded
Fluency	Number of different ideas produced	1 point for each idea
Flexibility	Number of categories of ideas produced	1 point for each category
Originality (Item 1 to 3)	Uniqueness of the ideas produced, as compared to the whole sample	< 5% - 2 points Between 5% and 10% - 1 point >10% - 0 point
Originality (Item 4)	Uniqueness of the ideas produced, as compared to the whole sample	< 5% - 3 points Between 5% and 10% - 2 points >10% - 1 point

Administration of Test

Form A and Form B were administered to a sample of 206 students from the fifth graders in two urban primary schools in Kota Kinabalu, Sabah, with a broad ability student intake. The schools were categorized in the middle band in terms of science achievement. Science achievement was assessed by the school science teachers, using end-of-semester test marks. A total of 115 students (55.8%) were female and 91 students (44.2%) were male. All participants were 11 years old. They completed Form A one day in May 2014 and Form B in the following week.

Data Analysis

The quantitative data were collected through Form A and Form B test scores and then analysed using SPSS (Windows V. 19.0). Item analysis included the calculation of item discrimination, internal consistency, inter-scorer reliability, construct-related validity and face validity.

Results of Research

Item Discrimination

According to Matlock-Hetzel (1997), a discrimination coefficient (DisCo) has advantages if compared to discriminant indices. DisCo includes all the test scores in the computation instead only 54% of total in upper and lower group are taken in computing discriminant indices. In order to compute the discrimination coefficient, the total test scores of students were ranked ascendingly. Then, the DisCo was calculated using formula for test items as suggested in Kiamanesh (2002) and Jandaghi (2010), i.e.

$$DisCo_i = \frac{s_i(upper) - s_i(lower)}{n_g * S_i}$$

Where

$s_i(upper) s_i(upper)$= Score for upper group in item ii.

$s_i(lower) s_i(lower)$= Score for lower group in item ii.

$n_g n_g$ = Number of student in one group.

$S_i S_i$ = Total score of item ii.

Table 2. Discrimination coefficient of test items.

	Discrimination Coefficient	
Item No.	**Form A**	**Form B**
Q1	0.29	0.32
Q2	0.29	*0.18
Q3	0.29	0.21
Q4	*0.19	0.21

The computed coefficient of item discrimination is indicated in Table 2. All items in Form A show a discrimination coefficient of 0.29, while the range is within 0.21 and 0.32 in Form B. Ebel's (1972) guidelines on discrimination indexes for item analysis, place items with index greater than 0.20 as an acceptable range. The coefficient indicates most of the items are acceptable and only two items are marginal (marked * in table) which are subject to improvement.

Internal Consistency

The Cronbach's Alpha of internal consistency was computed at the first stage to determine the degree to which the items on the same test, measure the same construct in order to produce a consistent result. The alpha coefficients based on 206 respondents were 0.77 in Form A and 0.68 in Form B. As a rule of thumb, an instrument which has reliabilities of 0.70 or higher will suffice in basic educational research (Nunnally, 1978: 245). Therefore, the alpha values of the instrument are considered to have acceptable reliability in the measure context.

The corrected item-total correlation and Cronbach's Alpha values, after each item is deleted, were also calculated as presented in Table 3. All the corrected item-total correlation and cronbach's alpha values are less than that if the item is not removed. It can be concluded that the overall of the result indicates all the items consistently contribute similar result and probing the same domain.

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 62, 2014

Table 3. Corrected item-total correlation & Cronbach's alpha if item deleted.

	Form A		Form B	
	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q1	0.432	0.764	0.360	0.663
Q2	0.428	0.787	0.086	0.738
Q3	0.494	0.758	0.339	0.664
Q4	0.213	0.777	0.418	0.656

Secondly, the inter-item correlations for Forms A and B were then calculated using Pearson product-moment correlation (Table 4). There were weak positive correlations in Form A ($p<0.01$) between item 1 and 3 ($r=0.228$), item 2 and 3 ($r=0.152$) and between item 1 and 4 ($r=0.220$), while for Form B there were weak positive correlations between item 1-4 ($r=0.333$), and between item 3-4 (0.209). There were also weak negative correlations between item 1-2 (-0.188), and item 2-4 (-0.250). Overall, there was a significant, weak, positive correlation between test items.

However, the result for inter-trait dimension correlation in each item of Form A and B (Table 5) did show a strong relationship, with total values for Form A ranging from 0.832 to 0.953 (item 1), 0.870 to 0.958 (item 2), 0.876 to 0.937 (item 3), and 0.998 (item 4). The total values in Form B range from 0.737 to 0.949 (item 1), 0.930 to 0.974 (item 2), 0.907 to 0.979 (item 3), and 0.975 to 0.990 (item 4). All conclusions are made at a confidence level of 99%.

Table 4. Correlation coefficients between items.

		Form A[a]				Form B[b]			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Correlation	Q1	1.000	0.034	0.228	0.220	1.000	-0.188	0.035	0.333
	Q2	0.034	1.000	0.152	-0.058	-0.188	1.000	0.032	-0.250
	Q3	0.228	0.152	1.000	0.071	0.035	0.032	1.000	0.209
	Q4	0.220	-0.058	0.071	1.000	0.333	-0.250	0.209	1.000
Sig. (1-tailed)	Q1	-	0.314	0.000	0.001	-	0.003	0.309	0.000
	Q2	0.314	-	0.015	0.205	0.003	-	0.326	0.000
	Q3	0.000	0.015	-	0.156	0.309	0.326	-	0.001
	Q4	0.001	0.205	0.156	-	0.000	0.000	0.001	-
Determinant = 0.876									
Determinant = 0.779									

Table 5. Correlation coefficients between Trait dimension in each item of Form A and Form B.

Inter-item	Trait dimension	Item 1				Item 2				Item 3				Item 4		
		F	FX	O	Total	F	FX	O	Total	F	FX	O	Total	FX	O	Total
A	F	1.000				1.000				1.000						
	FX	0.991	1.000			0.902	1.000			0.876	1.000			1.000		
	O	0.629	0.599	1.000		0.723	0.630	1.000		0.736	0.637	1.000		0.991	1.000	
	Total	0.953	0.941	0.832	1.000	0.958	0.910	0.870	1.000	0.937	0.876	0.911	1.000	0.998	0.998	1.000
B	F	1.000				1.000				1.000						
	FX	0.987	1.000			0.973	1.000			0.881	1.000			1.000		
	O	0.492	0.495	1.000		0.831	0.759	1.000		0.937	0.814	1.000		0.935	1.000	
	Total	0.949	0.949	0.737	1.000	0.974	0.942	0.930	1.000	0.979	0.907	0.978	1.000	0.975	0.990	1.000

Inter-Scorer Reliability

Due to the subjectivity of scoring criteria, it was necessary for an independent person to reliably interpret based on the same scoring system. The scores for 20 students' work were rated independently by a science teacher who did not engage with the instrument development and the researcher. The Pearson product-moment correlation coefficients between the two sets of scores for fluency and flexibility in Form A and Form B are presented in Table 6. The originality index was not included as it would be scored using the frequencies and percentages of the whole sample. Correlations between scores vary from 0.652 to 1.000. Most of the scorer agreement indicates strong correlation with an r-value greater than 0.8, except for item 4 in Form B. Hence, results suggest that the scoring procedure was sufficiently objective.

Table 6. Scorer agreement (n = 20 answer sheets for each question).

Item	Pearson product-moment correlation coefficients			
	Form A		Form B	
	Fluency	Flexibility	Fluency	Flexibility
1	0.870	0.809	0.878	0.918
2	0.972	0.931	0.979	0.960
3	0.912	0.862	0.984	0.976
4		1.000		0.652

Construct-related Validity

Factor Analysis

To validate a newly developed creativity test, its factorial validity must be tested. Factorial validity is a construct-related validity through factorial analysis of the test score. To determine whether further factorial analysis was appropriate, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Barlett's Test of Sphericity were run (Table 7). According to Kaiser (1974), values of KMO that greater than 0.5 are acceptable. Form A obtained value of 0.528, whereas Form B obtained 0.553. These values are in the 0.5 and 0.7 (considered moderate). It can be concluded, that the correlation patterns for the four items are relatively compact and able to yield distinct, reliable factors through factor analysis. Further, this result is supported by Barlett's test that shows significant values less than 0.001. Thus, it was found that factor analysis is appropriate to be carried out.

Table 7. Kaiser-Meyer-Olkin (KMO) and Barlett's Test.

	KMO Measure of Sampling Adequacy	Bartlett's Test of Sphericity		
		Approx. Chi-Square	df	Sig.
Form A	0.53	26.8	6	0.00
Form B	0.55	50.6	6	0.00

After the factor analysis was executed, the factorial analysis with principal components obtained two factors in Form A and Form B. Table 8 shows that the percentage of variance is explained 34.3% by factor 1 and 27.6% by factor 2 in Form A, a total of 61.9% of the variance is explained by these two factors, while the percentage of variance is explained by 38.9% and 26.2% in Form B, both factors are explained by 65.0% of the total variance.

In-depth analysis on the factor loading of each item (Table 9) implies item 1 (0.760) and 2 (0.732) have strong influence on factor 1, as the item 3 (0.800) and 4 (0.661) have greater load on factor 2 in Form A. The result shows, that factor 1 is influenced strongly by item 1 (0.714) and 3 (0.699) in Form B, and item 4 loaded greatly (0.930) on factor 2. From table 9, item 2 of Form B shows a negative value (-0.706) of factor loading which means that this item measures an opposite dimension of what it intends to measure (scientific creativity). In order to produce a good scientific creativity test looking at the domain of creativity, it is better to obtain unifactor and similar significant load for all items on one factor in the analysis (Hu & Adey, 2002). The result reveals that scientific creativity test has moderate construct-related validity. However, the analysis outcomes inferred that there are two factors that could affect the creativity test. All of the items in fact measure students' scientific creativity, especially items 1, 2 and 3. Hence, the first factor has high possibility on 'scientific creativity' while the second factor would be discussed further in Discussion.

Table 8. Factor extraction.

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
A-Q1	1.371	34.277	34.277	1.371	34.277	34.277	1.294	32.349	32.349
A-Q2	1.104	27.595	61.872	1.104	27.595	61.872	1.181	29.523	61.872
A-Q3	0.811	20.283	82.156						
A-Q4	0.714	17.844	100.000						
B-Q1	1.554	38.862	38.862	1.554	38.862	38.862	1.496	37.409	37.409
B-Q2	1.047	26.166	65.028	1.047	26.166	65.028	1.105	27.619	65.028
B-Q3	0.801	20.016	85.044						
B-Q4	0.598	14.956	100.000						
Extraction Method: Principal Component Analysis.									

Table 9. Factor loading of each item.

	Form A		Form B	
	Factor 1	Factor 2	Factor 1	Factor 2
Q1	0.760	-0.198	0.714	-
Q2	0.732	0.252	-0.706	0.251
Q3	-0.201	0.800	0.699	0.413
Q4	0.375	0.661	-	0.930
Extraction Method: Principal Component Analysis.				
Rotation Method: Varimax with Kaiser Normalization.				
Rotation converged in 3 iterations.				

Face Validity

There is also a need to determine the face validity in order to ensure that the items in the test actually measure scientific creativity. A total of eight senior science teachers from eight primary schools in Kota Kinabalu and Tuaran, Sabah were asked to respond to the question, 'which items in the test can measure scientific creativity of primary school students?' The result suggested a high degree of face validity among primary science teachers as they all agreed the eight items shall be retained in the scientific creativity test with some corrections on item wording.

Discussion

This study presented empirical work to validate a preliminary scientific creativity test. A parallel four-item test guided with Scientific Creativity Structure Model (SCSM) was developed for measuring scientific creativity among the fifth grade students. The resulting analysis of item discrimination, internal consistency, inter-scorer reliability, construct-related validity and face validity indicate different levels of the test's power and suitability.

Discrimination index obtained by each item of Form A and Form B has reached the threshold of an acceptable range. However, the item 4 in Form A and item 2 in Form B fall in the marginal discrimination index. This implies that these two items are not sufficiently statistically rigourous to discriminate scientific creativity between high creative thinkers and low creative thinker. There is a need to review these two items to develop better discrimination indexes.

Overall internal consistency of both tests suggested that they were developed with high reliability, as well as inter-scorer reliability. However, inter-item correlations showed a weak positive, but significant correlation between the items, and indicated a strong relation between the three trait dimensions in each item. As a whole, these results inferred that both Form A and B measure the three trait dimensions of fluency, flexibility, and originality in the scientific creativity test. Nevertheless, both Form A and B could fairly measure the product and process dimension of SCSM.

Results from inter-item correlation indicated that two factors exist in Form A and Form B, indicating that another factor could be measured by the same test apart from scientific creativity. First factor loaded significantly by the first two-items of Form A and first three-items of Form B, reflecting that these items measure scientific creativity as defined in the objective. Other than that, the second item in Form B was found to load negatively on factor 1. This can be

PROBLEMS
OF EDUCATION
IN THE 21st CENTURY
Volume 62, 2014

interpreted as highly creative thinkers potentially scoring lower than low creative thinkers. The second item in Form B requested students to suggest as many scientific words, they can relate to 'microorganism', a low creative thinker who was a high achiever in Biology would score well on this item. Most of the students might respond directly to the stimuli (microorganism) from what they used to know instead of contributing something novel about 'microorganism.' When the second item in Form A (magnet) and Form B (microorganism) was studied, Form A gained sufficient load on testing 'scientific creativity'. Therefore, it is strongly recommended to use stimuli which are related to real-life situations that students encountered in the environment, home, and school when revising this item in future.

Item 3 measured the student's scientific imagination to explain the phenomenon. However, students' responses to the statement of 'The sun has lost its light' were mostly related to the impact this would have on living things instead of using science knowledge to explain the phenomenon itself imaginatively. This item 3 of Form A has strong indication on the second factor, that is, creativity in general.

Item 4 in both tests were designed to measure students' creative ability in science problem solving. However, item 4 measured how well the students know the problem, method and the solution. Students used their own visual-spatial thinking to decide the different possible methods to divide a square into four equal pieces of the shape and to create as many different symbols from given five or less equal length stick. Thus, item 4 in both tests provided students with opportunities to highly demonstrate their creative problem-solving abilities in the domain of spatial analytical. Students with high scores in spatial analytical tests might not have high creative potential in science since general creativity should not represent the creativity in a specific field. As Musil and Ondrusek (1982) claimed that specific divergent tests have to be designed in order to improve the prediction of specific types of creativity. To conclude, item 4 of both tests have strong indication on the second factor that is spatial analytical thinking.

Sinha and Singh (1987) believed that a good test in scientific creativity should measure the novelty or originality in the abilities that the creative scientist would need. They thought that "properly loaded factors of novelty, flexibility, and fluency applied to different processes involved in scientific method of problem-solving appear to be a good measure of scientific creativity" (Sinha and Singh, 1987: 4). Thus, it is suggested that test items measuring science problem solving items should be revised to measure abilities that characterize scientists, such as analysing, synthesizing, and hypothesizing. For example, students can be asked to use the given hands-on materials to craft a simple scientific device or construction that helps in overcoming a problem in the environment.

Conclusion and Future Work

This study reports a preliminary attempt to develop and validate a parallel equivalent form of scientific creativity test for the fifth grade students. The scientific creativity test had high internal consistency, inter-scorer reliability and face validity. Both Form A and Form B of the test had an acceptable discrimination index range. The test showed a weak positive, but significant correlation between the items in product and process dimensions, but a strong correlation between the three trait dimensions of SCSM. Test items on science problem solving have a strong indication loading on spatial analytical thinking. Analysis suggests that items 1 and 2 of Form A, items 1 and 3 of Form B are acceptable in assessing the scientific creativity of the fifth grade students. However, Items 3 and 4 of Form A, items 2 and 4 of Form B require further work to improve their reliability and validity.

The researchers believe that the implications of this study are also relevant to science education researchers who are interested in the development of scientific creativity test for the fifth grade students. The study suggests further work to improve the validity of the test, as well as item discrimination index, and involve scientific method of problem-solving on science problem solving items. Other than that, considerable work can be done to diagnose the validity of a test such as determining its parallel form reliability, correlation with the other well-established scientific creativity test, criterion validity and experimental validity.

Acknowledgements

The research reported in this study was supported by the Sabah University of Malaysia, under Grant No. RAG0018-SSI-2012. Any opinions, viewpoints, findings, conclusions, suggestions, or recommendations are expressed by the authors and do not necessarily reflect the views of the Sabah University of Malaysia.

References

Amabile, T. M. (1983). *The social psychology of creativity.* New York: Springer-Verlag.

American Association for the Advancement of Science (AAAS) (2003). *Benchmarks for science literacy: Project 2061*. New York: Oxford University Press.

Aktamıs, H., Şahin-Pekmez, E., Can, B. T., & Ergin, O. (2005). *Developing Scientific Creativity Test.* University of Dokuz Eylul. Retrieved 2/10/2014, from http://www.clab.edc.uoc.gr/2nd/pdf/58.pdf

Craft, A. (2000). *Teaching creativity: philosophy and practice.* New York: Routledge.

Ebel, R. L. (1972). *Essentials of educational measurement* (1st Ed.). New Jersey: Prentice Hall.

Engineers' Council for Professional Development. (1947). *Canons of ethics for engineers*. New York: Engineers' Council for Professional Development.

Guilford, J. P. (1959). *Personality*. New York: McGraw-Hill.

Guilford, J. P. (1967). *The nature of human intelligence.* New York, NY: McGraw Hill.

Hu, W.P., & Adey, P. (2002). A scientific creativity test for secondary school students*. International Journal of Science Education, 24* (4), 389–403.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Jandaghi, G. (2010). Assessment of validity, reliability and difficulty indices for teacher-built physics exam questions in first year high school. *Educational Research and Review, 5* (11), 651-654.

Johnston, J. (2005). *Exploring primary science and technology education: Early exploration in science*. New York: McGraw-Hill International.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36.

Kiamanesh, A. R. (2002). *Assessment and evaluation in physics*. Tehran, Iran: Ministry of Education Publications.

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, *18* (1), 3-14. http://dx.doi.org/10.1207/s15326934crj1801_2

Kozma, R. B. (2010). *ICT policies and educational transformation*. Paris: UNESCO

Lubart, T. I. (1994). Creativity. In: R. J. Sternberg (Ed.). *Thinking and Problem Solving*, 289-332. London: Academic Press.

Matlock-Hetzel, S. (1997). *Basic concepts in item and test analysis.* Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX, January 23-25, 1997. Retrieved 1/10/2014, from http://files.eric.ed.gov/fulltext/ED406441.pdf

Mohamed, A. (2006). *Investigating the Scientific Creativity of Fifth-Grade Students.* Umi: Proquest Information and Learning Company

Musil, M., & Ondrusek, D. (1982). Relationships between internal and external criteria of scientific creativity. *Studia Psychologica*, *24* (1), 55-62.

Nunnally, J. C. (1978). *Psychometric theory* (2nd Ed.). New York: McGraw-Hill.

Ozbey, S., & Alisinanoglu, F. (2008). Identifying the general ideas, attitudes and expectations pertaining to science activities of the teachers employed in preschool Education. *Journal of Turkish Science Education*, *5* (2), 82-94.

Rawat, T. C. (2010). A study to examine fluency component of scientific creative talent of elementary stage students of Himachal Pradesh with respect to area, type of school and gender. *International Transactions in Humanities and Social Sciences, 2* (2), 152-161.

Rosen, Y., & Mosharraf, M. (2014). *Online performance assessment of creativity skills: findings from international pilot study.* Singapore: International Association for Educational Assessment.

Sahin-Pekmez, E., Aktamis, H., & Can, B. T. (2009). Exploring scientific creativity of 7th grade students. *Journal of Qafqaz University*, 204-214.
Sefertzi, E. (2000). *Creativity*. Report produced for the EC funded project INNOREGIO: dissemination of innovation and knowledge management techniques. Retrieved 2/10/2014, from http://www.adi.pt/docs/innoregio_creativity-en.pdf
Simonton, D. K. (2012). *Assessing scientific creativity: conceptual analyses of assessment complexities*. Commissioned paper, The Science of Science and Innovation Policy Conference, National Academy of Sciences.
Sinha, A. K., & Singh, C. (1987). Measurement of scientific creativity. *Indian Journal of Psychometry & Education, 18* (1), 1-13.
Torrance, E. P. (1974). *Torrance tests of creative thinking*. Lexington, MA: Personnel Press.
Smith, S. M., Gerkens, D. R., Shah, J. J., & Vargas-Hernandez, N. (2003). *Empirical studies of creative cognition in idea generation.* Paper presented at the Kellogg Team and Group Research Center's Conference on Creativity and Innovation in Organizations James L. Allen Center, Northwestern University, June 27-28, 2003. Retrieved 3/10/2014, from https://www.tamu.edu/faculty/stevesmith/SmithCreativity/Smith_Gerkens_Shah_Vargas-Hernandez.pdf
Shukla, J. P., & Sharma, V. P. (1986). Manual for Verbal Test on Scientific Creativity. National Psychological Corporation, Agra.
Torrance, E. P. (1990), *Torrance test of creative thinking: manual for scoring and interpreting results*. Bensenville, IL: Scholastic Testing Service, Inc.
Wilson, E. O. (1998). *Consilience: the unity of knowledge.* New York: Vintage Book.
Wolfinger, D. W. (2000). *Science in the elementary and middle school.* NY: Addison Wesley Longman, Inc.

Advised by Saleh A. Alabdulkareem, King Saud University, Saudi Arabia

Received: *November 08, 2014*　　Accepted: *November 17, 2014*

Nyet Moi Siew	PhD, Senior Lecturer, Faculty of Psychology and Education, Sabah University of Malaysia, Jalan UMS, 88400 Kota Kinabalu, Sabah Malaysia. E-mail: snyemoi@yahoo.com
Chin Lu Chong	Master in Science Education, Alumni, Sabah University of Malaysia, Kota Kinabalu, Sabah, Malaysia. E-mail: ev1lyn87@hotmail.com
Kim On Chin	PhD, Senior Lecturer, Center of Excellence in Semantic Agents, Block A, Faculty of Computing and Informatics, Sabah University of Malaysia, Jalan UMS, 88400, Kota Kinabalu, Malaysia. E-mail: kimonchin@gmail.com Website: https://sites.google.com/site/kimonchinsite/