



Facilitating Torrance Test of Creative Thinking Use in Malaysian TVET Research: The Initial Step of Inter-rater Reliability Determination

Ahmad Rizal Madar¹, Chew Eng Sun², & Hashima Hamid^{3*}

^{1,2,3}Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA

DOI: <https://doi.org/10.30880/jtet.2019.11.01.013>

Received 29th August 2019; Accepted 30th October 2000; Available online 31st March 2019

Abstract: Creativity is a valuable 21st century attribute, vital to innovation, and is considered as an important aspect in technical and vocational education and training (TVET). This study is prompted by the lack of international test use in Malaysian TVET creativity-related research where the Torrance Tests of Creative Thinking (TTCT), translated into the Malay Language (ML) could fill in this gap. The determination of inter-rater reliability coefficients cover five norm-referenced dimensions or measures of figural creativity, viz. fluency, originality, elaboration, resistance to premature closure, abstractness of titles, and the overall average score for the TTCT (Form A). This study involves correlation research where subjective answers to TTCT (ML) by 35 prospective teachers from a Teacher Education Institute - Temenggong Ibrahim Campus, Johor Baharu were marked by 3 independent raters, and subsequently analysed for correlations. Data analyses of IRC were carried out using interclass correlation analysis found in the Statistical Package for Social Science (SPSS) software. Excellent IRC values were obtained as follows: Fluency (.99), Originality (.99), Elaboration (.90), Resistance to Premature Closure (.93), Abstractness of Titles (.98), and overall average score (.99), with $r \geq .9$. The IRC values obtained for the TTCT (ML) version are consistent with values obtained for the English (international) TTCT version, viz. r , from .78 to .99. In conclusion, the TTCT (ML) version is valid and reliable data collection instrument for research on creativity, innovation or higher-order thinking skills involving TVET respondents whose mother tongue is Malay.

Keywords: Torrance Tests of Creative Thinking; inter-rater reliability coefficient; interclass correlation; figural creativity

1. Introduction

One of the aims of the latest Malaysian educational master plan, viz. the Malaysian Education Development Blueprint (2013 -2025) is to fast track technical and vocational education and training (TVET) so as to produce highly skilled and adaptable workforce who are innovative, creative and competitive (Yusof, Roddin, Warman and Deraman, 2017). In technical and vocational education, there are specific subjects such as Design and Technology where creative and innovative thinking are needed to generate ideas and design products. Furthermore, all outcomes of academic subjects under the national curriculum are subject to assessments based on higher order thinking skills (HOTS), starting with the Form 3 national school assessments in 2014 (Yuen & Sivanandam, 2014, December 24). Rajendran (2013) classified HOTS as comprising the higher cognitive levels of applying, analysing, evaluating and creating, of which creating or creative thinking is the highest order thinking skill. As such, there are aspects and evidence which showed the importance of creativity in the field of technical and vocational education where research could be conducted.

*Corresponding author: hashima@uthm.edu.my
 2019 UTHM Publisher. All right reserved.
penerbit.uthm.edu.my/ojs/index.php/jtet

Creativity is considered as a resource (Runco, 2014) that leads to successful innovation (Shiu, 2014) which, according to Amiruddin and Masek (2014), simplifies the operation of any complex task for end-users. It has been classified as a 21st century skill which is important in many subjects and described in analogy form “as a cousin to innovation” (Shiu (2014)). In the context of technical and vocational education and training (TVET), Alias (2016) was underlined the importance of innovation for enhancing the quality of TVET curriculum generally that creativity provides the ideas that drives innovation, and it can be deduced that creativity also plays an important role in TVET (Hunter, 2013). Meanwhile, Chew, Abd Hamid and Madar (2017) highlighted that, creativity is a valuable human mental attribute that is an important aspect of technical and vocational education.

Hence, there has been an awareness of the importance of creativity or creative thinking skills and these skills have been incorporated by curriculum developers into every new Malaysian school curriculum since the 1980s up till the present Standard Curriculum for Primary and Secondary Schools (KSSR and KSSM) (Chew & Madar, 2016). This is in line with The Malaysia Education Blueprint 2015-2025 (Higher Education) incorporates elements to tackle the uncertainty of the Fourth Industrial Revolution (4IR) As such, the Higher Education Framework 4.0 has been established to address the issues and challenges of 4IR that focused on creativity and creative thinking (Edward, 2018)

In this era of globalization, there is still the issue of not using standard international tests for the assessment of creativity in Malaysia and every effort should not be spared to encourage their use. Chew and Madar (2016) conducted a documentary search on postgraduate theses involving creative thinking from electronic database at the Educational Planning and Research Unit of the Malaysian Ministry of Education, and found that only three out of 65 studies involved the use of international standardized instruments such as the Torrance Tests of Creative Thinking (TTCT) for validation or for assessing the creativity levels of respondents in research. Shiu (2014) noted that the TTCT (figural version) involves mostly drawing. Torrance (2018a) stated that the TTCT can be used in all cultures. In the Malaysian context, to test for creativity using a psychometric instrument, such as the TTCT, it has to be accurately translated into the Malay Language and its reliability coefficient determined, with an excellent level as the outcome before it is ready for use by researchers.

According to Brown, Irving and Keegan (2008), apart from validity, reliability is an important consideration in relation to assessment and the statistic that is frequently used in estimating reliability is the correlation statistic. Hence, this study focuses on the aspect of evaluation where a test instrument viz., the TTCT, has been translated into the Malay Language (ML) but without any current reliability references being available for all the five measures of Fluency, Originality, Elaboration, Resistance to Premature Closure, and Abstractness of Titles.

2. Related work

In Malaysia, TTCT has also been used to find out figural creativity and cognitive preferences among Malaysian undergraduates (Palaniappan,1998) and also in assessing secondary school students thinking and learning styles for effective teaching and learning (Chua, 2011) and also to study about assessment of creativity in Electrical Engineering (Afida Ayob et al, 2012).

2.1 Inter-rater Reliability

Fraenkel, Wallen, and Hyun (2012) defined reliability as the degree to which scores obtained with an instrument are consistent measures of whatever the instrument measures and the reliability coefficient as the index that expresses this consistency of scores. Brown, Irving and Keegan (2008) listed four types of comparison that are commonly used to estimate reliability, viz. time to time comparison (test-retest), assessment to assessment comparison (method error), marker to marker comparison (inter-rater) and item to total score comparison (internal).

Creswell (2012) attributed inter-rater reliability to the reliability associated with two or more scorers whereby the consistency of their scores for similarity are determined. Fraenkel, Wallen and Hyun (2012) suggested that for instruments that could be influenced by differences in administration or scoring or both, such as subjective evaluations, researchers are obliged to report the degree of scoring agreement.

According to Landers (2015), an interclass correlation can be used to estimate inter-rater reliability on quantitative data as it is highly flexible. Koo and Li (2016) stated that the interclass correlation coefficient (ICC) is often chosen as the reliability index in test-retest, intra-rater, and inter-rater reliability analysis. For scoring or rating on subjects or targets (rates) by the interclass correlation analysis which is based on the Generalizability Theory (G-Theory) offers three models of calculation viz. one way random effects or ICC (1), two-way random effects or ICC(2) and two-way mixed (ICC(3)). Categorization into any of these three models would involve looking into the different types of arrangements the inter-raters are allocated to score or mark on their rates (McGraw & Wong, 1996; Landers, 2015).

2.2 The Torrance Tests of Creative Thinking

In the assessment of creativity of individuals or groups of people using psychometric tests, the Torrance Tests of Creative Thinking (TTCT) stands out as a set of widely used and highly recommended tests in the field of education (Kim, 2006). It is a battery of standardized tests for a diversity of age groups from kindergarten students to adults (Torrance, 2018a) and has been translated into more than 35 languages at the turn of the 21st century (Millar, 2002). It has been re-normed a total of six times, four times before the turn of the century in 1974, 1984, 1990, and 1998 (Kim, 2006) and twice after that in 2007 and 2016 (Torrance, 2018a).

The TTCT is based on Guilford's (1975) premise that creative thinking involves an intellectual operation called divergent production. It emphasizes on the measures of fluency, originality, elaboration and flexibility. However, in the development of the TTCT instruments involving figural drawing, the flexibility component was replaced by Resistance to Premature Closure and Abstractness of Titles (Hebert, et al., 2002; Torrance 2018a). According to Torrance (2018b), fluency refers to the number of relevant or meaningful ideas that are produced. Originality is based on the statistical infrequency or 'unusualness' of the figure that is drawn. Elaboration is referred to as the attention to detail shown to the object of attention and its surroundings. Resistance to premature closure highlights the tendency of the respondent not to close an open figure too quickly. In choosing a suitable title for a drawn picture, synthesis and organization of thought as well as relevant abstractness are the criteria required to have a high score for abstractness of titles. The TTCT is a pencil and paper test that offers two approaches to the testing of creative thinking, viz. the TTCT Figural tests and the TTCT-Verbal tests. Testing is based on five norm-referenced measures and 13 criterion-referenced creative strengths. Each approach comes in two parallel forms viz. Form A and Form B which can be used as a pretest and a post-test respectively (Ststesting, 2018).

This study involves only the TTCT-Figural Form A which consists of three drawing activities which are assessed on the five norm-referenced measures of fluency, originality, elaboration, resistance to premature closure and abstractness of titles. Each activity involves figural drawing and title creation and has to be completed in 10 minutes. According to Kim (2006) and Torrance (2018b), Activity 1 invites the test taker to construct a picture based on a pear-shaped blot. The test taker is given the chance to elaborate on any part of the picture to make it look as complex as he pleases. In Activity 2, ten incomplete sets of partially drawn figures are given as stimuli. Each stimulus is used to generate the drawing of an object or picture. In Activity 3, many sets of twin lines are given as stimuli for drawing objects or pictures individually or in a combination of lines. In all activities, a suitable title is created by the test taker to describe the completed object or picture. The TTCT-Figural Form A, the TTCT - Streamlined Scoring Guide and the Norms- Technical Manual are given to raters to facilitate marking. Since 1984, a streamlined scoring system has been created to help in providing details and in simplifying scoring procedures (Torrance, 2018b).

3. Methodology

This study involves the correlational design (Creswell, 2014), where sets of marks from three raters are analysed for inter-rater reliability correlation which shows scoring agreement. Creswell (2014) suggested that the correlational design, which is a type of non-experimental quantitative research design, is suitable to measure the degree of association between two or more variables or sets of scores. According to Fraenkel, Wallen and Hyun (2012), researchers using subjective test instruments like the TTCT need to report the degree of scoring agreement involving inter-raters.

Three lecturers from a Malaysian teacher education institute, *Institut Pendidikan Guru-Kampus Temenggong Ibrahim Johor Baharu (IPG-KTIJB)* were chosen as test-markers or inter-raters. Inter-raters were chosen purposively to mark subjective test responses involving forms of drawings and creative titles from test takers. The respondents who took the TTCT tests in Malay comprise 35 prospective teachers from the Counselling and Guidance and the Design and Technology specialist options from IPG-KTIJB and are chosen by purposive sampling. Inter-raters were given two sessions of briefing and moderation before marking the responses independently. The selection of the inter-raters was based on the criteria of i) having had the experience of teaching the infusion of creative thinking in teacher education courses and ii) having authored journal articles related to creative thinking. Data is in the form of TTCT (ML) test scores of 35 respondents whose work were assessed according to the five measures of fluency, originality, elaboration, resistance to premature closure, abstractness of titles and their overall average scores. As such, for each measure, three sets of marks from 35 respondents were obtained and analyzed for correlation. The interclass correlation application feature found in the Statistical Package for Social Science (SPSS) were used to analyze the data which is criterion referenced. In the context of this study where all the inter-raters mark all the respondents' test scripts, the interclass correlation coefficient (ICC) is calculated by the SPSS software.

4. Results

This study uses the interclass correlation analysis from the SPSS package to show the degree of correlation between the inter-raters' scores for the measures of figural creativity in TTCT viz. fluency, originality, elaboration, resistance to premature closure and abstractness of titles and their overall average score. Raw scores from the TTCT are used in the data analysis. Descriptive statistics in Table 1 shows the mean scores for the 5 measures of figural creativity and their

corresponding standard deviations achieved by the 3 inter-raters. In Table 1, mean scores from all the three inter-raters are presented together with their respective standard deviations.

Table 1- Items reliability for the corresponding competency constructs

No	Measures	N	Inter-rater 1		Inter-rater 2		Inter-rater 3	
			Mean	S.D	Mean	S.D	Mean	S.D
1.	Fluency	35	23.09	8.35	22.83	8.22	23.60	8.03
2.	Originality	35	17.46	6.95	16.97	6.75	17.89	6.67
3.	Elaboration	35	4.51	1.42	4.23	1.37	4.94	1.49
4.	Resistance to Premature Closure	35	11.49	3.94	10.57	3.56	11.37	4.12
5.	Abstractness of Titles	35	4.09	3.28	3.46	2.94	4.00	2.96

Using the median of the mean scores from the three inter-raters as representative score for each measure in Table 1, the five measures of figural creativity are ranked in descending order as follows: Fluency (23.09) > Originality (mean = 17.46) > Resistance to premature closure (11.37) > Elaboration (4.51) > Abstractness of titles (4.00)

The highest score (mean=23.09; SD=8.35) comes from the component measure of fluency, which is followed by originality (mean=17.46; SD=6.95), resistance to premature closure (mean=11.37; SD=4.12). The 2 lowest scoring measure with single digit scores are *elaboration* (mean=4.51; SD=1.42) and *abstractness of titles* (mean=4.00; SD=2.96). Inter-rater 1 has the highest scores in 2 measures, viz. *resistance to premature closure and abstractness of titles* while Inter-rater 3 has the highest score in 3 measures, viz. *fluency, originality and elaboration*. Inter-rater 2 is the strictest marker with lowest scores in all the measures. Table 2 displays the reference for interpreting the level of reliability based on the reliability coefficients obtained. According to Koo and Li (2016), four categories of interpretation can be identified based on the reliability coefficient, *r*, namely, poor, moderate, good and excellent.

Table 2 - Interpretation of ICC as a reliability index according to levels of reliability (Koo & Li, 2016)

Level of reliability	Poor	Moderate	Good	Excellent
Range of <i>r</i> values	≤ .5	Between .5 and .75	Between .75 and .9	≥ .9

Fraenkel, Wallen and Hyun (2102) claimed that the desired or highly acceptable correlation for inter-rater correlation coefficient is at least .90 or agreement of at least 80% among scorers.

4.1 Fluency

Table 3 shows the values of ICC for the component measure of fluency.

Table 3 - ICC for the measure of Fluency

	Interclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p- value.
Single Measures	.974 ^b	.954	.986	119.203	34	68	.000
Average Measures	.991 ^c	.984	.995	119.203	34	68	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A interclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 3, the reliability coefficient of correlation for Fluency (as represented by the average measures of interclass correlation) is found to be 0.991. According to Koo and Li (2016), a reliability coefficient value of equal or greater than .90 indicates a correlation among raters that is excellent. This implies that for Fluency, the marks between the three raters are not significantly different at $p = .05$.

4.2 Originality

Table 4 shows the data about the ICC for the measure of Originality.

Table 4 - ICC for the measure of Originality

	Interclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p-value
Single Measures	.963 ^b	.934	.980	87.285	34	68	.000
Average Measures	.987 ^c	.977	.993	87.285	34	68	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A interclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 4, the reliability coefficient of correlation for Originality is found to be 0.987. According to Koo and Li (2016), this is an indication that the correlation between the 3 sets of marks for Originality is at an excellent level. This implies that for Originality, the marks between the 3 raters are not significantly different at $p = .05$.

4.3 Elaboration

Table 5 shows the data about the ICC for the measure of Elaboration.

Table 5 - ICC for the measure of Elaboration

	Interclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p-value
Single Measures	.749 ^b	.571	.862	12.424	34	68	.000
Average Measures	.900 ^c	.800	.949	12.424	34	68	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A interclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 5, the IRC for Elaboration is found to be 0.90. According to Koo and Li (2016), this indicates that the correlation among the three sets of marks for Elaboration is at an excellent level. This implies that for Elaboration, the marks between the three raters are not significantly different at $p = .05$.

4.4 Resistance to Premature Closure

The ICC for measure of Resistance to Premature Closure is shown in Table 6.

Table 6 - Interclass correlation coefficient for the measure Resistance to Premature Closure

	Interclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p-value
Single Measures	.749 ^b	.571	.862	12.424	34	68	.000
Average Measures	.900 ^c	.800	.949	12.424	34	68	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A interclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 6, the reliability coefficient of correlation as indicated by for Resistance to Premature Closure is found to be .930. According to Koo and Li (2016), this indicates that the correlation among the three sets of marks is at an excellent level. This implies that for the measure of Resistance to Premature Closure, the marks between the three raters are not significantly different at $p = .05$.

4.5 Abstractness of Titles

Table 7 below describes the data of ICC for the measure of Abstractness of Titles.

Table 7 - Interclass correlation coefficient for the measure Abstractness of Titles

	Interclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p-value
Single Measures	.928 ^b	.869	.962	46.132	4	8	.00
Average Measures	.975 ^c	.952	.987	46.13	4	8	.00

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A interclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 7, the reliability coefficient of correlation for Abstractness of Titles is found to be $r=.975$. According to Koo and Li (2016), this indicates that the correlation among the three sets of marks for Abstractness of Titles is at an excellent level and not significantly different at $p = .05$

4.6 Overall average score for five measures

Table 8 describes the data of ICC for measure Overall Average Score for the 5 measures.

Table 8 - Interclass correlation coefficient for the Overall Average Score

	Interclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	p-value
Single Measures	.963 ^b	.916	.983	109.90	34	68	.00
Average Measures	.987 ^c	.970	.994	109.90	34	68	.00

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type A interclass correlation coefficients using an absolute agreement definition.

b. The estimator is the same, whether the interaction effect is present or not.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

In Table 8 the reliability coefficient of correlation for the Overall Average Score for five measures of figural creativity is found to be $r=.987$. According to Koo and Li (2016), this indicates that the correlation among the three sets of marks for overall average score is at an excellent level and not significantly different at $p = .05$. This indicates that a great consistency in marking among the three inter-raters.

4.7 Summary on the ICC of All Measures of Figural Creativity and the Overall Average Score

This section shows the summary of all the reliability the data about the interclass correlation coefficient for overall average of five measure of figural creativity as shown in Table 9.

Table 9 - Summary of ICC for all the five measures of figural creativity and the Overall Average Score

Measures	Fluency	Originality	Elaboration	Resistance to Premature Closure	Abstractness of Titles	Overall Average Score
Inter-raters' reliability coefficient	.991	.987	.900	.930	.975	.987
Level of reliability	excellent	excellent	excellent	excellent	excellent	excellent

In Table 9, a summary of inter-raters' reliability coefficient for all the 5 measures, all correlation coefficient values fall in a range between .90 and .99. Hence, according to Koo and Li (2016), it has excellent reliability and all the 3 set of marks are not significantly different at $p = .05$.

5. Discussion

5.1 The reliability coefficient for all the measures of figural creativity in Form A of the TTCT (Malay Language version)

Inter-rater values for the Malay Language version of the TTCT in this study are higher than r values obtained from an earlier study from Tan (1992) which range from .73 to .99. Inter-rater coefficient values from this study are also in agreement with values obtained for the international (English Language version) of the TTCT, which span the same range of values except for college level assessment for Resistance to Premature Closure ($r = .78$) (Torrance, 2018a). With the findings of this study where the value of the reliability coefficient of correlation is equal to or greater than .90, the agreement among inter-raters is found to be at an excellent level. This indicates that the TTCT as a test instrument has high consistency when it is used for creativity measurement using Malay Language in the study.

5.2 The high level of agreement among inter-raters in the marks given for the 5 measures of figural creativity

According to Fraenkel, Wallen and Hyun (2012), training which include explaining, discussing and moderating on the scoring process is repeated until inter-raters reached an excellent level of agreement. In this study, a high level of agreement among inter-raters was reached after 1 session of briefing and two sessions of moderation as the selected inter-raters are not only practitioners on the creative teaching and learning process in their institutes but are also creativity researchers.

5.3 The ranking of the five measures of Figural Creativity based on the reliability coefficient for the measures of Figural Creativity in TTCT (ML version)

Based on the values of inter-raters' correlation coefficient found for the five measures of figural creativity, their ranking in descending orders are given in Table 10.

Table 10 - Ranking of creativity measures according to inter-rater reliability coefficient values in descending order

	Fluency	Originality	Abstractness of titles	Resistance to premature closure	Elaboration
Measures	.991	.987	.975	.930	.90
Ranking in descending order					

The scoring for responses to the TTCT tests is based on the TTCT streamline scoring guide (Torrance, 2018b). According to the ranking shown in Table 10, inter-raters have the highest agreement for fluency and originality but lowest for elaboration. The high degree of agreement among inter-raters for fluency and originality is expected as both measures employ marking criteria that are more objective compared to the other three measures. As an example, when doing the count for fluency, inter-raters just have to decide and include each drawing by determining that the pictorial responses do not have duplicates or are not relevant to the stimulus. For originality, answers found in a list that disqualifies a drawn picture as not original are left out of the count. Elaboration is expected to have the least agreement among inter-raters as the marks given to embellishments made by the respondents to the main drawing is highly subjective and left more to the interpretation of each marker (Torrance, 2018b).

5.4 Impact of the Determination of the Inter-Raters' Reliability Coefficients for the Malay Language Version of the Torrance Tests of Creative Thinking for research in TVET

Research on creativity or creative thinking, innovation and higher order thinking skills among TVET respondents may require a creativity test instrument. With the translation of the TTCT in Malay Language and update of the correlation coefficient in the form of inter-rater reliability coefficients, Malaysian researchers have the option of carrying out research using a standardized international creativity test instrument, the results of which can be compared on an international scale. Research can be conducted to determine the current creativity levels of TVET respondents, or in conjunction with an interventionist method. For example, the effectiveness of a creativity enhancement module or

instructional aid on Malaysian polytechnic students where the TTCT (ML) version can be used as a pre-test (Form A) and post-test (Form B).

6. Conclusion

The inter-rater reliability coefficients for the five measures or dimensions of figural creativity, and the overall average score in Form A of the TTCT (ML version) instrument were determined by using the interclass correlation analysis found in the SPSS package. The updated values for all measures are at a desired or excellent level for inter-rater reliability coefficients. It is also in agreement with the correlation values obtained for the international (English) version of the TTCT. Comparison with values obtained from a much earlier study using the TTCT (ML version), viz from .73 to .99, shows that the reliability coefficient values obtained from this study is at a higher range of equal to and above .90. As supported by theory, with sufficient training and the correct choice of inter-raters, the inter-raters involved in scoring the TTCT (ML version) tests could reach a high level of agreement. In conclusion, this study has provided an update on the reliability coefficients for the five measures and overall creativity of TTCT (ML version) and the overall average score and this constitutes a current standard of reference for reliability for this test instrument. It would be of benefit to local researchers who choose to assess creativity by using test measurements rather than using surveys based on perceptions of their respondents' own creative ability, and also for respondents who are Malay speakers. As creativity or creative thinking, innovation and higher-order thinking skills are currently important focuses of research in Malaysian education, the TTCT (ML) instrument with a known reliability coefficient will help and facilitate TVET researchers to advance research in these areas.

Acknowledgement

This research was conducted to fulfill the requirement of the PPG Research Grant with Vot. No. 032 provided through the Office for Research, Innovation, Commercialization and Consultancy Management, Universiti Tun Hussein Onn Malaysia. The authors would like to thank the Malaysian Ministry of Education and all relevant agencies for making possible the completion of this research and publication of this article.

References

- Ayob, A., Hussain, A., Mustafa, M. M., Abdul Majid, R.(2012). Assessment of Creativity in Electrical Engineering *Procedia - Social and Behavioral Sciences* (60), 463 – 467
- Alias, M. (2016, October). Enhancing the quality of TVET through innovations. Keynote Address given at the 5th *World Congress on Technical and Vocational Education and Training (WoCTVET 2016)* in KSL Hotel and Resort, Johor Bahru, Malaysia.
- Amiruddin, M.H., & Masek, A. (2014). *Inovasi dalam teknologi pendidikan: Isu dan cabaran*. Parit Raja, Malaysia: Penerbit UTHM.
- Brown, G., Irving, E., & Keegan, P. (2008). *An introduction to educational assessment, and evaluation*. 2nd Edn. North Shore, NZ: Pearson Prentice Hall
- Chew, E.S., Abd Hamid, M.A., & Madar, A.R. (2017). Conceptual framework for designing and developing a creativity enhancement module in education incorporating indigenous perspectives. *Pertanika Journal of Social Sciences and Humanities*, 25(5), 67-82.
- Chew, E.S., & Madar, A.R. (2016). Pola dan trenda dalam penyelidikan berkaitan dengan kreativiti atau Pemikir kreatif pada peringkat pasca ijazah di Malaysia. [Patterns and trends in research related to creativity or creative thinking at postgraduate level in Malaysia]. *Jurnal Pemikir Pendidikan*. 7, pp.73-88.
- Chua Y. P. & Mohd Don. Z (2011). A Grounded theory for Research Implementation in Schools. Proceedings of International Conference on Applied Social Sciences ICASS 2011. Volume 2, pp413- 417, 19-21 March, Changsha, China.
- Creswell, J.W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative Research*. 4th Edn. Boston MA: Pearson
- Creswell, J.W. (2014). *Research design: Qualitative, quantitative , & mixed methods approaches*.4th Edn. SAGE Publications Asia-Pacific: Singapore.
- Edward, O.T. (2018). Creative, Critical Thinking is key. News Straits Time Online. Accessed on 26 November, 2018.
- Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). *How to design and evaluate research in education*. 8th Edn. New York, NY: McGraw-Hill.
- Guilford, J.P. (1975). Creativity: A quarter century of progress. In Irving, A.T. & Getzels, J.W. (Eds.) *Perspectives in Creativity*. Chicago: Aldine Publishing Company. pp. 37-59.

- Hebert, T.P., Crammond, B., Neumeister, K.L.S., Millar, G. & Silvian, A.F. (2002). *E. Paul Torrance: His Life, Accomplishments, and Legacy*. Storrs: The University of Connecticut, The National Center on the Gifted and Talented.
- Hunter, S. (2013). *Out Think: How innovative leaders drive exceptional outcomes*. Mississauga, Canada: Jossey Bess.
- Kim, K.H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creative Research Journal*, 18(1), 3-14.
- Koo, T.K., & Li, M.Y. (2016). A guideline of selecting and reporting interclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 14 (2), 155-163. doi:10.1016/j.jcm. 2016. 02. 012
- Landers, R. (2015). Computing interclass correlations (ICC) as estimates of inter-rater reliability in SPSS: *The Winnower* 2. doi: 10.15200/winn.143518.81744.
- McGraw, K.O. & Wong, S.P. (1996). Forming inferences about some interclass correlation coefficients. *Psychological Methods*. 1 (1), 30-43.
- Millar, G.W. (2002). *The Torrance kids at Mid-Life*. Westport: Ablex.
- Palaniappan, A.K. (1998). Figural Creativity and Cognitive Preference Among Malaysian Undergraduate Students. *The Journal of Psychology*, 132 (4), 381-388.
- Ststesting. (2018). Gifted Education: Torrance Tests of Creative Thinking. Retrieved from <http://www.ststesting.com/ngifted.html>
- Rajendran, N.S. (2013). *Teaching & Acquiring Higher-Order Thinking Skills: Theory & Practice*. Tanjong Malim: Universiti Pendidikan Sultan Idris.
- Runco, M.A. (2014). The psych economic perspective on creativity and innovation. In Shiu, E. (Ed.). *Creativity research: An inter-disciplinary and multi-disciplinary research handbook*, i. New York, NY: Routledge.
- Shiu, E. (2014). *Creativity research: An inter-disciplinary and multi-disciplinary research handbook*, 1st Edition. New York, NY: Routledge.
- Tan, A. (1992). *A validation study of the Torrance Figural Test of Creative Thinking* [Abstract]. Unpublished Masters Thesis, Universiti Malaya, Malaysia.
- Torrance, E.P. (2018a). *The Torrance Tests of Creative Thinking: Norms-Technical Manual for figural Forms A and B*. Bensenville, IL: Scholastic Testing Service, Inc.
- Torrance, E.P.(2018b). *The Torrance Tests of Creative Thinking: Streamline Scoring Guide for figural Forms A and B*. Bensenville, IL: Scholastic Testing Service, Inc.
- Yuen, M. & Sivanandam, H. (2014, Dec 24). NUTP to submit proposals to improve PT3 system. *The Star*, p.2
- Yusof, Y., Roddin, R., Warman, S., & Deraman, N.A. (2017). *Aspirasi Pelan Pembangunan Pendidikan Malaysia (PPPM) 2013-2025 terhadap hala tuju kecemerlangan bidang Pendidikan Teknik and Vokasional. (The aspiration of the Malaysian Education Development Blueprint towards excellence in the field of technical and vocational education. Putrajaya: Kementerian Pendidikan Malaysia.)*