LOOK ACADEMIC PUBLISHERS
OPEN ACCESS
INTERNATIONAL JOURNAL OF ENVIRONMENTAL & SCIENCE EDUCATION
e-ISSN: 1306-3065
2019, Vol. 14, No. 4, 139-149

Assessing the Validity and Reliability of the Future Thinking Test using Rasch Measurement Model

Nyet Moi Siew [1*], Mohammad Syafiq Abd Rahman [1]

[1] *Universiti Malaysia Sabah, MALAYSIA*

* CORRESPONDENCE: snyetmoi@yahoo.com

ABSTRACT

This research aimed to assess the validity and reliability of a developed Future Thinking Test (FTT) using the Rasch measurement model. The FTT consisted of 12 items concerning knowledge of physical science which were scored using Rating Scale Model. Items were developed based on the five future thinking constructs: i) Understanding the current situation; ii) Identifying trends; iii) Analyzing relevant drivers; iv) Developing scenarios of possible and probable futures, and v) Selecting with justification the preferable futures. Rasch analyses were conducted on data from a sample of 66 tenth graders (36 females and 30 males, aged 16 years old) from two rural secondary schools in the Tawau district, Sabah, Malaysia. Overall, the FTT was found to have good reliability with a Cronbach's alpha value (KR-20) of 0.69. Results also showed that FTT has an excellent item reliability and moderate high item separation value of 0.97 and 5.92 respectively. FTT also has a sufficient good person reliability and person separation value of 0.67 and 1.41 respectively. The sub-constructs are individually unidimensional and the future thinking with five sub-constructs are also unidimensional. Meanwhile, the validity of the FTT instrument was appropriately established through the item fit, except for one item. Results showed that 11 items were appropriate while one item needed to be discarded to increase the validity and reliability of FTT. This research indicates that FTT has an acceptable validity and reliability for measuring the five constructs of future thinking among tenth graders in rural secondary schools.

Keywords: futures thinking test, Rasch measurement model, tenth graders

INTRODUCTION

Infusing future thinking into the curriculum is a recent movement in educational reforms in many developed countries. Future thinking is perceived as one of the skills that students need to master in order to cope with whatever kind of circumstances in the future (Haapala, 2002). Bishop and Hines (2012) also stressed that the ability to illustrate future entities from various aspects created a think tank community capable of analyzing future possibilities, predicting future needs and choosing the desired future.

There is little evidence that instruments are being developed to assess students' future thinking. Past research show that future thinking was tested in psychology and psychometrics either via oral test or questionnaires (Addis, Hach, & Tippett, 2016; Brown, Root, Romano, Chang, Bryant, & Hirst, 2013; Maccallum & Bryant, 2011; Raffard, Esposito, Boulenger, & Van der Linden, 2013). There were studies which assessed the differences in future thinking between adolescents and adults (Cole, Morrison & Conway, 2013) and the public's retrospective future preferences (Noblet, Anderson & Teisl, 2015). However, very few instruments were developed using written tests to measure future thinking among students. Carter, Creedy and Sidebotham (2015) stressed that the measurement of future thinking in the classroom is very important

Article History: Received 7 July 2019 ◆ Revised 18 July 2019 ◆ Accepted 19 July 2019

Table 1. Achievement of Average Subject Grade in Sciences in the Tawau District

School Category	Total of schools	Average Subject Grade (ASG)	Grade
Urban areas	7	3.29	B to B+
Rural areas	14	4.67	C+ to B
Difference in ASG		**1.41**	

Source: Tawau Assessment & Examination Unit (2019)

to identify the development of students' cognitive ability. Thus, a future thinking test which is specifically targeted at the level of school classrooms is required to assess the future thinking of students. The present study addressed this concern by developing a future thinking test for tenth graders and assessing its validity and reliability using Rasch measurement model.

Statistics show that a total of 156 schools, which are about 73% of secondary schools in Sabah, Malaysia are categorized as rural schools (Sabah State Education Department, 2017). In terms of infrastructure, most rural schools in Sabah have no access to good teaching and learning resources, computers and science laboratories (SEDIA, 2011). These limited opportunities and facilities have created a gap in the academic achievement between rural and urban schools in Sabah. In the Tawau District, for instance, a difference of 1.41 in the Average Subject Grade was recorded between the urban and the rural schools in science subjects in a national examination of the Malaysian Certificate of Education (**Table 1**).

Sustainability is a topic in the Grade 10 Science syllabus offered in the Malaysian secondary school system. Sustainability necessitates learning from the past, exploring the present, thinking about the future, and developing solutions that are adaptable and resilient (Warren, Archambault, & Foley, 2014). As such, in-depth studies need to be carried out to investigate the level of future thinking among rural school students in science subjects. Through acquirement of future thinking, rural school students can master science knowledge, understand and evaluate it from all aspects of the future. It is pertinent for a future thinking test to be developed and rooted in the local environment, society and culture based on the rural school context. Hence, a future thinking test which is precisely focused on the context of rural schools was proposed to assess the future thinking of tenth graders in Sabah.

LITERATURE REVIEW

Future Thinking

Future thinking refers to the imagination of possible future events (D'Argembeau, Ortoleva, Jumentier, & Van der Linden, 2010) or one's ability to creatively imagine the limitless possibilities of hypothetical future scenarios (Fortunato & Furey, 2011). The cognitive process of future thinking involves episodic and semantic memories (Martin-Ordas, Atance, & Caza, 2014). Episodic memory provides the materials for individuals to imagine their future lives; semantic memory retrieves knowledge of the world providing individuals to imagine things that might happen in the future. Although the future is unpredictable, individuals can simulate potential future scenarios in their mind (Schacter, Addis, & Buckner, 2008; Szpunar, 2010).

On the other hand, Amara (1981) defined future thinking as the identification, discovery, analysis and evaluation of future possibilities, probabilities and requirements. Likewise, future thinking was described by Bishop and Hines (2012) as a way to think of future possibilities, future needs and desired future selections. In this study, future thinking refers to conceptual framework founded by Jones, Buntting, Hipkins, Conner, and Saunders (2012), which are i) understanding the current situation, ii) identifying key trends, iii) analysing relevant drivers, iv) developing scenarios of possible and probable future, and v) selecting with justification of the desired future.

The conceptual framework of future thinking test

A future thinking model (**Figure 1**) proposed by Jones *et al.* (2012) was used as a conceptual framework to guide the development of test items in the Future Thinking Test. Based on **Figure 1**, the construct of understanding the current situation is an attempt to explore the overall event, channeling science knowledge and linking context with individuals and social aspects. The construct of identifying key trends refers to the determination of pattern changes in events that can be observed nowadays or seen in the future and is caused by the changes of driver. The construct of analyzing relevant drivers is an analysis of the factors that cause changes, affect or give effect on something. The fourth construct, "developing scenarios of a possible and probable future", refers to efforts to make the future more evident in decision-making to produce new thoughts

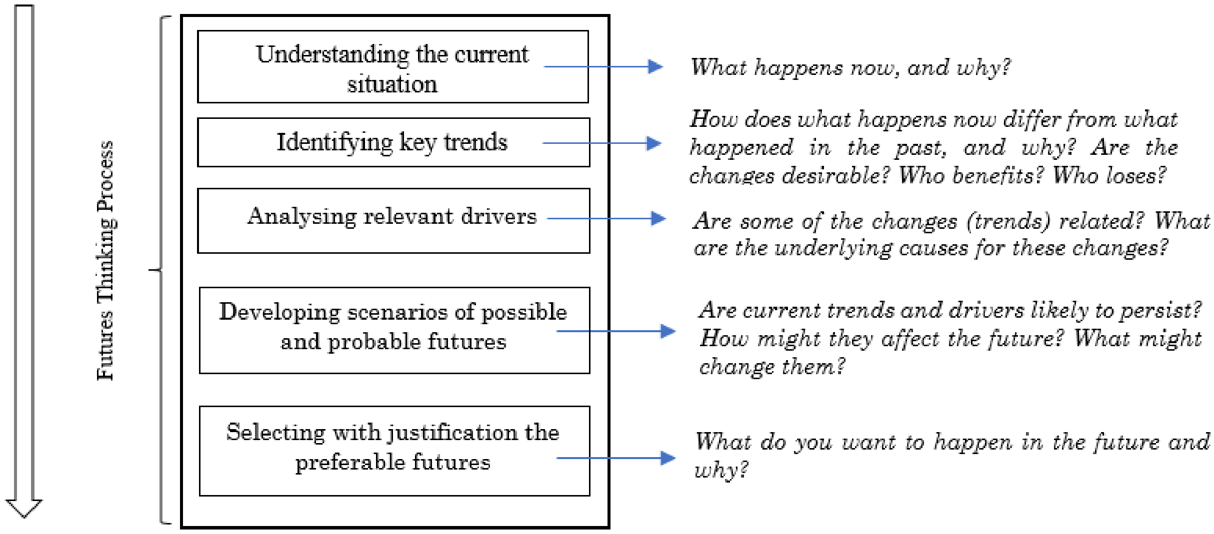

Figure 1. Conceptual framework of futures thinking by Jones *et al.* (2012)

and results, to learn how to think of all opportunities available to explore. The last construct, selecting with justification the preferable futures, deals with the desired hope(s), aspiration(s) and dreams in the future through exploration of existing opportunities.

This framework seeks to operationalize the interconnections between and among sustainability-related topics in the Science subject such as energy sources and environmental conservation. This framework provides a landscape in which teachers can situate science knowledge, sustainability and pedagogy as well as craft meaningful lessons that foster future thinking.

METHODOLOGY OF RESEARCH

Development of FTT

The Future Thinking Test (FTT) is an open-ended scenario-based test supported with photos illustrating family houses of the past and the present. The scenario entailed the changes in family houses which occurred in the students' village. The added description was given as follows to trigger students' future thinking:

Family houses have changed a lot over the last hundred years. Changes in family houses are not specific to its size but the change in internal and external structure as well. Are these changes desirable? How do you imagine a family house will be like in the future?

The students then answered the FTT which consisted of 5 items and 7 sub-items. The constructs and related items assessed were: i) Understanding the current situation (Item 1(a), 1(b)), ii); Identifying trends (Item 2(a), 2(b)); iii) Analyzing the relevant driver (Item 3(a), 3(b), 3(c)); iv) Selecting with justification the preferable futures (Item 4(a), 4(b)), and v) Selecting with justification of desired future requirements (Item 5(a)(i), 5a(ii), 5(b)) (**Table 2**).

Table 2. Scoring criteria for future thinking test

Item	Components	Scoring criteria	Score awarded
1	**Understanding the current situation** a) List three (3) features of the family house you know at the present	Number of relevant features	• '0' point = irrelevant features • '1' point = 1 relevant feature • '2' points = 2 relevant features • '3' points = 3 relevant features
	b) Give three (3) reasons why the given features are such.	Number of relevant reasons	• '0' point = irrelevant reasons • '1' point = 1 relevant reason • '2' points = 2 relevant reasons • '3' points = 3 relevant reasons
2	**Identifying key trends** a) Based on the pictures given, give three (3) differences between family houses of the past and the present.	Number of differences	• '0' point = irrelevant differences • '1' point = 1 relevant difference • '2' points = 2 relevant differences • '3' points = 3 relevant differences
	b) Give three (3) key changes (trends) that can be observed in houses today.	Number of key changes (trends)	• '0' point = irrelevant trends • '1' point = 1 relevant trend • '2' points = 2 relevant trends • '3' points = 3 relevant trends
3	**Analyzing relevant drivers** a) Give three (3) underlying causes (drivers) for these changes/trends	Number of underlying causes (drivers)	• '0' point = irrelevant drivers • '1' point = 1 relevant driver • '2' points = 2 relevant drivers • '3' points = 3 relevant drivers
	b) Give three (3) advantages of these drivers.	Number of advantages	• '0' point = irrelevant advantages • '1' point = 1 relevant advantage • '2' points = 2 relevant advantages • '3' points = 3 relevant advantages
	c) Give three (3) disadvantages of these drivers.	Number of disadvantages	• '0' point = irrelevant disadvantages • '1' point = 1 relevant disadvantages • '2' points = 2 relevant disadvantages • '3' points = 3 relevant disadvantages
4	**Developing scenarios of possible and probable futures** a) Will these trends and factors continue to affect the features of future houses? Give three (3) reasons.	Number of relevant reasons	• '0' point = irrelevant reasons • '1' point = 1 relevant reason • '2' points = 2 relevant reasons • '3' points = 3 relevant reasons
	b) Provide three (3) possibilities or requirements for the future house in line with the development of factors and the changing of current trends.	Number of possible and probable future house	• '0' point = irrelevant possibilities • '1' point = 1 relevant possibilities • '2' points = 2 relevant possibilities • '3' points = 3 relevant possibilities
5	**Selecting with justification the preferable futures** a) If you are given the opportunity to create a family house for community use for the next 50 years, what would you suggest? i) Sketch and label three (3) features that may be required in a future house the space provided. ii) Create one (1) name for your future house and state two (2) materials used.	Number of features of possible future house Name and materials	• '0' point = irrelevant sketches • '1' point = a rough sketch with 1 feature • '2' points = a fairly detailed sketch with 2 features • '3' points = a detailed sketch with 3 features • '0' point = irrelevant names and materials • '1' point = a rough name with 1 material • '2' points = a fairly unique name with 2 materials • '3' points = a unique name with 2 materials
	b) Provide three (3) justifications why the features are selected for the future house.	Number of relevant justifications	• '0' point = irrelevant justifications • '1' point = 1 relevant justification • '2' points = 2 relevant justifications • '3' points = 3 relevant justifications

Item 1(a) required the students to first understand the current scenario of a family house and list out its features. Item 1(b) required students to give reasons why the features were such as stated in Item 1(a). Item 2(a) required students to provide the differences between the old and present house based on the pictures given. Students were also instructed to give three major changes (trends) that could be observed in the present family house in item 2(b). Item 3(a) required the students to give the underlying causes (drivers) for those changes. Next, Item 3(b) and 3(c) expected students to answer three advantages and three disadvantages of these drivers respectively. Item 4(a) required the students to give reasons as to whether these trends and factors would continue to influence the production of a family house in the future. Students were also required to give three possibilities of future family houses in line with the development of factors and changes in the current family house in item 4 (b). In Item 5(a)(i), students were to produce and label the sketches of a future model if they were given the opportunity to create a family house for community use in the next 70 years. Item 5(a)(ii) required the students to name their future family house and state the materials used while Item 5(a)(iii) necessitated the students to provide three justifications for the features selected for their future family house.

A time frame of 60 minutes was specified to complete the test whereas students were subjected to answer each item and sub-item in 5 minutes.

Scoring procedure

The scoring rubrics suggested by Ho, Wang and Cheng (2013) and the Rating Scale Model proposed by Bond and Fox (2015) were used as a guide to evaluate the five constructs of future thinking. The quality of the students' responses was ranked from "0" (Level 0: the lowest level) to "3" (Level 3: the highest level). **Table 2** shows the scoring criteria used for assessing future thinking.

Rasch Measurement Model

In the construction of items which aim to measure a person's behaviour, Wright and Stone (1979) suggested that it is important for instrument developers to be aware of the crucial task of creating high quality items. Items which are high in quality deal better with the validity and reliability of the instrument.

An instrument is said to have a high validity if the instrument can measure what is supposed to be measured (Creswell, 2014; Pallant, 2005). Reliability is the extent to which a research instrument produces consistent results if it is used in the same situation on repeated occasions. According to Bond and Fox (2007), Rasch measurement model is an effective model in diagnosing and analyzing the validity and reliability of the items contained in a test. This is also voiced by Summintono and Widhiarso (2014) that Rasch measurement model can identify a person's validity in determining if the items used can produce consistent results in different groups of subjects. According to Boone, Yale, and Staver (2014), the person validity is salient to ensure the data collected by the researcher is accurate. Therefore, in this study, the Rasch analysis was used to determine the validity and reliability of the developed future thinking test.

Research Purpose and Research Questions

The purpose of this research was to develop a future thinking test (FTT) for tenth graders relating to science subjects and to assess the validity and reliability of the instrument using Rasch measurement model. Accordingly, the research questions guiding this research were:

What is the validity and reliability of the future thinking test in the constructs of:

1) Understanding the current situation,

2) Identifying the trends,

3) Analyzing relevant drivers,

4) Synthesising the future possibilities or needs, and

5) Selecting with justification the preferable futures?

Administration of the FTT

The Future Thinking Test was administered in two secondary rural schools in October 2018. The two selected schools are located in a rural area on the East Coast of Sabah, Malaysia. School A was about 80 kilometers whereas school B was 152 kilometers from Tawau. The participants consisted of 66 tenth graders, with 30 males (45.5%) and 36 females (54.5%), aged 16 years old. According to Cooper and Schindler (2011), the sample size ranging from 25 to 100 is sufficient to produce consistent and accurate findings.

Prior to administering the test, the social-scientific approach with Bloom's future thinking maps were employed as an explicit approach to teaching future thinking. The science lessons were designed to teach future thinking, physical science (energy, electric, heat and light), and sustainability. The respondents transformed the five constructs of future thinking into a future thinking map to explain the changes of the family house. Thus, students could gain benefit from the explicitness of the thinking maps that guide, direct, and incite their future thinking. The intervention consisted of six lessons of 2 hours each, conducted within 6 weeks.

Table 3. Fit Indices for Item Fit

Statistics	Fit Indices
Outfit mean square values (MNSQ)	0.50 – 1.50
Outfit z-standardized values (ZSTD).	-2.00 – 2.00
Point Measure Correlation (PTMEA-CORR)	0.40 – 0.85

Source: Boone *et al.* (2014)

Data Analysis

The obtained data were analyzed using WINSTEP version 3.73. There are three fit indices criteria sets for establishing the reliability of the future thinking test (FTT) which are Cronbach's alpha KR-20, item and person reliability, and item and person separation. According to Bond and Fox (2007), the test is considered to have a high reliability, is strong and acceptable if the test has a separation value higher than 2.0, and the item and person reliability value higher than 0.8. Linacre (2007) stated that a separation value that is higher than 2.0 is considered ideal. As for the reliability of Cronbach's alpha value, Bond and Fox (2007) published that the value between 0.71-0.99 is considered to be high and at its best value.

On the other hand, the validity of the FTT was established based on the misfit order of the items. According to Sumintono and Widhiarso (2015), item fit can inform the researcher whether the item functions normally in performing the supposed measurements, as well as to assess the suitability of the item. Boone, Staver and Yale (2014) and Bond and Fox (2015) suggested three criteria to be used for assessing the item fit, which are, Outfit Mean Square Values (MNSQ), Outfit Z-Standardized Values (ZSTD), and Point Measure Correlation (PTMEA-CORR). According to Bond and Fox (2007), Outfit MNSQ informs the researcher about the suitability of the item in measuring the construct, while PTMEA-CORR informs the extent to which the development of the constructs achieved its goals. A positive PTMEA-CORR value indicates that the item measured the construct to be measured, while a negative PTMEA-CORR value indicates otherwise. On the other hand, ZSTD are *t*-tests of the hypothesis which can inform the researcher whether the data perfectly fits the model. According to Sumintono and Widhiarso (2015), any item that fails to fulfill these three criteria (**Table 3**) necessitates improvement or changes to ensure the quality and suitability of the item.

Other than that, the assessment on the Principal Component Analysis (PCA) was carried out to test the dimensionality of the future thinking test. The standardized residual variance in the PCA to test the dimensionality (unidimensionlaity/multidimensionality) of an instrument can be identified easily by using Rasch measurement model compared to other models (Sick, 2011; Tennat & Pallant, 2006; Wright, 1999). On top of that, Jusoh, Amlus, Osman, and Abidin (2014) also argued that the analysis through PCA in determining the dimensionality of an instrument is important since the items which are suitable (fit) with the Rasch measurement model and fall within the unidimensionality category indicates that the instrument has a high content validity (Wright, 1996; Sick, 2011). According to Reckase (1979), the dimensionality of an instrument can be determined based on these two important elements which are 'raw variance explained by measures' and 'unexplained variance in the 1st- 5th contrast'. Reckase (1979) suggested that the raw variance explained by measures should exceed 20%. However, Rasch argued that a value which is more than 40% was the best indicator for the unidimensionality of a test. Meanwhile, for the unexplained variance in the 1st- 5th contrast, the ideal value is below 15% (Sumintono & Widhiarso, 2014).

RESEARCH RESULTS

Misfit Order of the Items (Item Fit)

Table 4 shows the misfit order of the items which informs about the suitability of the items measuring the constructs of the FTT instrument. The misfit order of the items presented was based on the value of Outfit MNSQ, Outfit ZSTD and PTMEA-CORR. The bold figures imply that the items failed to fulfill the criteria suggested by Boone *et al.* (2014). The PTMEA-CORR ranged from 0.26 - 0.75 for items 1b, 2a, 2b, 3a, 3b, 3c, 4a, 4b, 5a (i) and 5b. A positive index indicates that the items measure the constructs to be measured (Bond & Fox, 2007). However, the PTMEA-CORR for item 1a and 5a (ii) were found to be -.02 and -.16. Bond and Fox stated that the negative value of PTMEA-CORR suggests that the item does not have the ability to measure the constructs. Accordingly, these two items need to be refined or discarded. It is also discovered that only item 5a (ii) has an Outfit MNSQ value above the 1.5 logit which is 2.36, while the rest of the items have an Outfit MNSQ value within the range between 0.5 to 1.5 logit. Jailani (2011) argued that the Outfit MNSQ value provided more robust value compared to the Infit MNSQ value in determining the suitability of the item

Table 4. Misfit Order of the Items in FTT

Item	MEASURE	Outfit MNSQ (0.50-1.50)	Outfit ZSTD (-2.0-2.0)	PTMEA-CORR (0.40-0.85)
5b	1.52	.87	-.6	.50
5a(ii)	1.02	**2.36**	**5.6**	**-.16**
5a(i)	.97	.97	-.1	.26
4b	.71	.69	-2.1	.69
4a	.65	.92	-.4	.49
3c	.52	.82	-1.2	.66
3b	.00	.85	-1.0	.75
3a	-.39	.84	-1.0	.69
2b	-.62	.93	-.3	.62
2a	-.82	.82	-1.0	.39
1b	-1.15	1.16	.9	.51
1a	-2.41	1.45	1.4	**-.02**

Table 5. Principal Component Analysis (PCA)

STANDARDIZED RESIDUAL Variance (in Eigenvalue units)					
		---- Empirical ----	Modeled		
Total raw variance in observations	=	24.9	100.0%		100.0%
Raw variance explained by measures	=	12.9	**51.7%**		50.4%
Raw variance explained by persons	=	3.8	15.2%		14.8%
Raw Variance explained by items	=	9.1	36.6%		35.6%
Raw unexplained variance (total)	=	12.0	48.3%	100.0%	49.6%
Unexplained variance in 1st contrast	=	2.6	**10.4%**	21.5%	
Unexplained variance in 2nd contrast	=	2.4	**9.7%**	20.0%	
Unexplained variance in 3rd contrast	=	1.2	**4.9%**	10.1&	
Unexplained variance in 4th contrast	=	1.1	**4.4%**	9.2%	
Unexplained variance in 5th contrast	=	.9	**3.7%**	7.7%	

for the measurement of the construct. Thus, item 5a (ii) should be dropped according to the expert's view as it may be misleading and difficult to be answered by the respondent (Linacre, 2007).

Mohd Jailani (2011) also asserted that items that exceed the 1.5 logit of the Outfit MNSQ value should be dropped to ensure that the item matches the constructs to be measured. In addition to the Outfit MNSQ value, the Outfit ZSTD value also should be within the range of -2 to +2, however, the Outfit ZSTD value can be neglected if the Outfit MNSQ value has already met with the conditions of the suitability of the item (Bond & Fox, 2007; Linacre, 2007). Conclusively, the summary of Outfit MNSQ, Outfit ZSTD and PTMEA-CORR values is shown in **Table 4**.

Dimensionality

Based on **Table 5**, the Principal Component Analysis (PCA) (which is based on the value of raw variance explained by measures) is 51.7%. This finding indicates that the FTT instruments are accepted and categorized as unidimensionality instruments (Reckase, 1979) and have high content validity evidence (Jusoh, Amlus, Osman, Dba & Abidin, 2014). Whereas the unexplained variance in the 1st - 5th contrast has value less than 15% which denote that items in the FTT instrument are strong and valid in measuring the five constructs of future thinking (Sumintono & Widhiarso, 2014).

Reliability of Cronbach Alpha (KR-20) Value

The findings demonstrate that the Cronbach's alpha (KR-20) value which is 0.69 shows that the FTT instrument has a very moderate high reliability of internal consistency (Sumintono & Widhiarso, 2014). Thus, this evidence shows that the FTT instrument can be used in the actual research.

Item Reliability and Separation

Rasch measurement model also diagnosed and analyzed the FTT instrument thoroughly in terms of reliability and separation of items and person (research subject). **Table 6** shows that the item reliability is 0.97 while the item separation is 5.92. Based on the item reliability, the value of 0.97 indicates that the item has a high reliability (Bond & Fox, 2007; Linacre, 2007; Summintono & Widhiarso, 2014). Meanwhile, the value of item separation (5.92) in **Table 6** signal that the items in the FTT instrument can be broken down

Table 6. Reliability and separation of item and person

	Item	Person
Reliability	0.97	0.67
Separation	5.92	1.41

Table 7. Cronbach's Alpha (KR-20), Item Reliability, Item Separation, Person Reliability and Person Separation

	Rasch Measurement Model	Future Thinking Test (FTT)	Interpretation
Cronbach Alpha (KR-20)	>0.60	0.69	Good
Item Reliability	>0.80	0.97	Very High
Item Separation	>2.0	5.92	Very High
Person Reliability	>0.80	0.67	Sufficient
Person Separation	>2.0	1.41	Moderately Good

Table 8. Summary of the Remained, Purified, and Dropped Items

Construct	Item	Item Remained	Item Purified	Item Removed
Understanding the Current Situation	1(a), 1(b)	1(b)	1(a)	None
Identifying the Key Trends	2(a), 2(b)	2(a), 2(b)	None	None
Analyzing the Relevant Drivers	3(a), 3(b), 3(c)	3(a), 3(b), 3(c)	None	None
Synthesizing the Scenarios of Possible and Probable Futures	4(a), 4(b)	4(a), 4(b)	None	None
Selecting with Justifications for Preferable Futures	5(a)(i), 5a(ii), 5(b)	5(a)(i), 5(b)	None	5a(ii)
Total	**12**	**10**	**1**	**1**

into several levels of difficulties (Bond & Fox, 2007). Linacre (2007) also argued that a good value for item separation should be greater than 2.0.

Person Reliability and Separation

Table 6 also shows that the value of person reliability is 0.67 while the value of person separation is 1.41. Based on the Rasch measurement model, the value of 0.67 for the person reliability is sufficient and acceptable (Summintono & Widhiarso, 2014). Summintono and Widhiarso (2014) also argued that the person reliability value of 0.67 is sufficient to produce similar findings for different groups. However, Bond and Fox (2007) stated that the person reliability should be higher than 0.8 to produce consistent responses for different groups.

Besides that, the person separation value (1.41) signify that the FTT instrument is in a good, moderate category and acceptable. The separation value of 1.41 suggests that the research subject can be grouped into two groups of student's capabilities on the item difficulty levels. Linacre (2007) and Summintono and Widhiarso (2014) stated that the value of person separation on the item difficulty levels is appropriate to produce heterogeneous research subjects if the person separation value is more than 2.0.

Altogether, the reliability of the Cronbach's alpha (KR-20) value, item separation and item reliability value, person separation and person reliability value are presented in **Table 7**.

Summary of Findings

The findings established that the validity and reliability of the FTT instrument fulfilled the standard index and conditions suggested by the Rasch measurement model in measuring the construct of students' future thinking. Based on the item fit analysis through PTMEA-CORR, Outfit ZSTD and Outfit MNSQ value, it was discovered that item 1(a) needs to be purified while item 5a(ii) needs to be removed. Item 1(a) will be purified by referring to the views and opinions from the experts in the test development. This is to ensure the improvement of validity and reliability of the items in the FTT instrument as well as to increase its quality. Conversely, items 1(b), 2(a), 2(b), 3(a), 3(b), 3(c), 4(a), 4(b), 4(c), 5(a) (i) and 5(b) which showed a positive PTMEA-CORR value and fall within the logit range, will be retained as these items are appropriate and suitable to measure students' future thinking. The summary of the items in the FTT instrument (after the analysis using Rasch measurement model) is presented in **Table 8**.

DISCUSSION

The Rasch measurement model has enabled the researcher to detect problematic items, so that they can be improved or excluded to increase the validity and reliability of the instrument to measure the constructs of future thinking. Overall, the analysis discovered that an item in the FTT instrument needed purification while another had to be dropped. It is necessary to remove or purify items that are incompatible with the Rasch model in order to better the validity and reliability of the FTT instrument. Other than that, the analysis from the Principal Component Analysis (PCA) in the Rasch model detected that the sub-constructs of FTT instrument were individually unidimensional and that the FTT with the five sub-constructs was also unidimensional. This signals that the FTT is able to measure the five constructs of students' future thinking.

Meanwhile, in terms of reliability, the FTT instrument presented a good internal consistency based on the Cronbach's alpha (KR-20) value, and has a very high item reliability. Thus, it can be said that the FTT instrument is a reliable instrument for measuring students' future thinking constructs, particularly among the rural tenth grade students in Tawau, Sabah. Besides that, it was discovered that the FTT instrument has a very high value of item separation which show that the instrument has a greater spread of item (Klooster, Taal & Laar, 2008). Md Yunos, Ibrahim Mukhtar, Alias, Lee, Tee, Rubani, Hamid, Yunus, Sulaiman, and Sumarwati (2017) argued that an item separation value which exceeds 2.00 indicates that the items in the instrument are strongly accepted. Whereas, Krishnan and Idris (2004) stated that the item separation value which exceeds 1.00 indicates that the items in the instrument have enough spread.

On top of that, the analysis from Rasch model also revealed that the FTT instrument has a fairly good person reliability and has a moderately good person separation. The person separation value of 1.41 which exceeds 1.00 connotes that the students in the study can be well distinguished (Gracia, 2005). Meanwhile, Krishnan and Idris (2014) stated that the person separation must be over 1.00 to ensure that students are measured across the continuum. However, some authors (Md Yunos *et al.* 2017; Khamis & Che Yahya, 2015) considered that the separation value of less than 2.00 to be low. Thus, purifying or removing the problematic items in the FTT instrument may improve the reliability as well as the separation index value.

CONCLUSION

In context, it is proven through the examination on the validity and reliability of the FTT instrument that it is a suitable and reliable instrument that was applied by the researcher to assess students' future thinking. Besides that, establishing the validity and reliability of the FTT instrument in measuring future thinking constructs was a crucial step taken by the researcher to ensure that the actual research could be executed effectively in the future. More importantly, the implication from the Rasch measurement model in establishing the validity and reliability of the FTT instrument has been instrumental to the researcher in enhancing the quality of the instrument for the findings (from using the FTT instrument) to be valid and reliable.

ACKNOWLEDGEMENTS

The research was supported by the Universiti Malaysia Sabah (UMS), Malaysia under Grant No. FRG0462-2017. Any opinions, viewpoints, findings, conclusions, suggests, or recommendations expressed are the authors and do not necessarily reflect the views of the Universiti Malaysia Sabah, Malaysia.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Nyet Moi Siew – Faculty of Psychology and Education, Universiti Malaysia Sabah, Malaysia.

Mohammad Syafiq Abd Rahman – Faculty of Psychology and Education, Universiti Malaysia Sabah, Malaysia.

REFERENCES

Addis, D. R., Hach, S., & Tippett, L. J. (2016). Do strategic processes contribute to the specificity of future. Simulation in depression? *British Journal of Clinical Psychology*, *55*(2), 167-186. https://doi.org/10.1111/bjc.12103

Amara, R. (1981). The futures field: Searching for definition and boundaries. *Futures, 15*(2), 25-29.

Bishop, P. C., & Hines, A. (2012). *Teaching about the Future*. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137020703

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch Model: *Fundamental Measurement in the Human Sciences* (2nd eds.). Mahwah, NJ: LEA.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd eds.). New York: Routledge. https://doi.org/10.4324/9781315814698

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

Brown, A. D., Root, J. C., Romano, T. A., Chang, L. J., Bryant, R. A., & Hirst, W. (2013). Overgeneralized autobiographicalmemory and future thinking in combat veterans with posttraumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(1), 129-134. https://doi.org/10.1016/j.jbtep.2011.11.004

Carter, A. G., Creedy, D. K., & Sidebotham, M. (2015). Evaluation of tools used to measure critical thinking development in nursing and midwifery undergraduate students: A systematic review. *Nurse Education Today*, *35*(7), 864–874. https://doi.org/10.1016/j.nedt.2015.02.023

Cole, S. N., Morrison, C. M., & Conway, M. A. (2013). Episodic future thinking: Linking neuropsychological performance with episodic detail in young and old adults. *Quarterly Journal of Experimental Psychology*, *66*(9), 1687–1706. https://doi.org/10.1080/17470218.2012.758157

Cooper, D., R., & Schindler, P., S. (2011). *Bussines research methods*. (11th eds.). New York: McGraw-Hill/Irwin.

Creswell, J., W. (2014). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research* (4th eds). USA: Pearson Education Limited.

D'Argembeau, A., Ortoleva, C., Jumentier, S., & Van Der Linden, M. (2010). Component processes underlying future thinking. *Memory and Cognition*, *38*(6), 809–819. https://doi.org/10.3758/MC.38.6.809

Furtunoto, V.J., & Furey, J.T. (2011). The theory of MindTime: The relationships between Future, Past and Present thinking and psychological well-being and distress. *Personality and individual Differences*, *50*(1), 20-24. https://doi.org/10.1016/j.paid.2010.08.014

Gracia, S. (2005). Analyzing CSR Implementation with the Rasch Model" (2005). *Faculty Publications*. 271. Retrieved from https://digitalcommons.ric.edu/facultypublications/271

Haapala, A. (2002). Mitä on tulevaisuuskasvatus (*What is Future Education*?). In A. Haapala (Eds.), *Tulevaisuuskasvatus (The Future of Education)* (pp. 125-140). Jyväskylä: PS-kustannus.

Ho, H. C., Wang, C. C., & Cheng, Y. Y. (2013). Analysis of the scientific imagination process. *Thinking Skills and Creativity*, *10*, 68-78. https://doi.org/10.1016/j.tsc.2013.04.003

Jailani, K. M. (2011). *Manual Pengenalan Pengukuran Rasch and Winstep. Pengukuran dan Penilaian dalam Pendidikan (Introduction of Rasch and Winstep Measurement Introduction. Measurement and Evaluation in Education).* Fakulti Pendidikan Universiti Kebangsaan Malaysia.

Jones, A., Buntting, C., Hipkins, R., McKim, A., Conner, L., & Saunders, K. (2012). Developing students' futures thinking in science education. *Research in Science Education*, *42*(4), 687–708. https://doi.org/10.1007/s11165-011-9214-9

Jusoh, M. S., Amlus, M., H., Osman, A., & Abidin, R. (2014). Development and Validation of a Multidimensional Measure: Manufacturing Practices, Knowledge Transfer and Manufacturing Capabilities in Malaysian Manufacturing Companies. *Australian Journal of Basic and Applied Sciences*, *8*(12), 25-29.

Khamis, M. R., & Che Yahya, N. (2015). Does Law Enforcement Influence Compliance Behaviour of Business Zakat among SMEs? An Evidence via Rasch Measurement Model. *GJAT*, 5(1), 19-32. https://doi.org/10.7187/GJAT752015.05.01

Klooster, P. M., Taal., E., & Laar, M. A. F. J. (2008). Rasch Analysis of the Dutch Health Assessment Questionnaire Disability Index and the Health Assessment Questionnaire II in Patients with Rheumatoid Arthritis. *Arthritis & Rheumatism (Arthritis Care & Research), 59*(12), 1721–1728. https://doi.org/10.1002/art.24065

Krishnan, S., & Idris, N. (2014). Investigating Reliability and Validity for the Construct of Inferential Statistics. *International Journal of Learning, Teaching and Educational Research, 4*(1), 51-60.

Linacre JM. 2009. *A user's guide to FACETS Rasch-model computer programs, program manual 3.65.0.* Chicago, IL: Winsteps.com.

Maccallum, F., & Bryant, R. A. (2011). *Imagining the future in complicated grief. Depression and Anxiety*, *28*(8), 658-665. https://doi.org/10.1002/da.20866

Martin-Ordas, G., Atance, C. M., Caza, J. S. (2014). How do episodic and semantic memory contribute to episodic foresight in young children? *Frontiers in Psychology*, *5*(732), 1-11. https://doi.org/10.3389/fpsyg.2014.00732

Md Yunos, J., Ibrahim Mukhtar, M., Alias, M., Lee, M. F., Tee, T.K., Rubani, S.N.K., Hamid, H., Yunus, F.A.N., Sulaiman, J., & Sumarwati, S. (2017). Validity of Vocational Pedagogy Constructs Using The Rasch Measurement Model. *Journal of Technical Education and Training*, *9*(2), 35-45.

Noblet, C. L., Anderson, M. W., & Teisl, M. F. (2015). Thinking past, thinking future: An empirical test of the effects of retrospective assessment on future preferences. *Ecological Economics*, *114*, 180-187. https://doi.org/10.1016/j.ecolecon.2015.04.002

Pallant, J. (2005). *SPSS Survival Manual-a Step by step Guide to Data Analysis Using SPSS fow Windows (Version 12).* NWS Australian: Allen and Unwin.

Raffard, S., Esposito, F., Boulenger, J. P., & Van der Linden, M. (2013). Impaired ability to imagine future pleasant events is associated with apathy in schizophrenia. *Psychiatry Research*, *209*(3), 393-400. https://doi.org/10.1016/j.psychres.2013.04.016

Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, *4*, 207-230. https://doi.org/10.2307/1164671

Sabah State Education Department (2017). Senarai Sekolah Rendah Kategori Luar Bandar Dan Pedalaman Di Negeri Sabah (*List of Rural and Inland Schools in Sabah*). Retrieved from http://www.ums.edu.my/upplbv2/files/SENARAI-SEKOLAH-RENDAH-KATEGORI-LUAR-BANDAR-DAN-PEDALAMAN-DI-NEGERI-SABAH.pdf

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic simulation of future events: Concepts, data and applications. *Annals of the New York Academy of Sciences*, *1124*, 39–60. https://doi.org/10.1196/annals.1440.001

SEDIA (2011). Sabah Economic Development and Investment Authority Blueprint. Retrieved from http://www.sedia.com.my/

Sick, J. (2011). Rasch Measurement and Factor Analysis. *Testing & Evaluation SIG Newsletter*, *15*(1), 15-17.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial (edisi revisi). Application of the Rasch Model for Social Sciences Research (revised edition).* Cimahi: Trim Komunikata Publishing House.

Szpunar, K. K. (2010). Episodic future thought. *Perspectives on Psychological Science*, *5*(2), 142–162. https://doi.org/10.1177/1745691610362350

Tennant, A. & Pallant, J. F. (2006). Unidimensionality matters (A tale of two Smiths?). *Rasch Measurement Transaction*, *20* (1), 1048-1051.

Warren, A., Archambault, L., & Foley, R. W. (2014). Sustainability Education Framework for Teachers: Developing sustainability literacy through futures, values, systems, and strategic thinking**.** *Journal of Sustainability Education*, *6*, 1-14.

Wright, B. D. (1999). Common sense for measurement. *Rasch Measurement Transactions*, *13*(3), 704.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA Press.