Development and Validation of an Instrument to Measure STEM Teachers' Instructional Preparedness

Nur Farhana Ramli[1*], Othman Talib[2], Siti Aishah Hassan[3] & Umi Kalthom Abdul Manaf[4]

[1*]Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
ramli.nurfarhana@gmail.com
[2]Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
otalib@upm.edu.my
[3]Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
siti_aishahh@upm.edu.my
[4]Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor
umizat90@upm.edu.my
*Corresponding Author
http://doi.org/10.24191/ajue.v16i3.11084

Abstract: STEM education curriculum is an effort mainly targeted for stakeholders to increase students' interest in STEM as well as to meet future workforce demand. Teachers' instructional preparations are an important part to the overall effectiveness of the teaching and learning process. This study focuses on the development and validation process of STEM Teachers' Instructional Preparedness Instrument (STEMTIP). The study comprises ten steps ranging from development to validation phase. Social constructivist theory, 5E instructional model and STEM teaching and learning approach model were used as the basis for the development of this instrument. There were 51 items initially generated using inductive and deductive approaches which include face validity, content validity, construct validity and criterion validity as were confirmed by the experts', and statistical results of Rasch model and regression analysis. The reliability analysis also demonstrated acceptable consistency for the instrument. After going through the validity process, a total of 41 items remained as final items for this instrument. Full details of reliability analysis, validation along with implication for practice are discussed.

1. Introduction

STEM is an acronym for Science, Technology, Engineering and Mathematics disciplines, integrated as a multidisciplinary approach. It began with the launch of the world's first artificial satellite, Sputnik by Russia in 1957 (U.S. Information Agency, 1959) which resulted in the advancement of science and technology especially in aerospace science. Hence, it sparked competition between countries such as the US, and the United Kingdom to be the leader in the field of science and technology.

Recently, STEM-related careers have become a sign of prestige around the world due to the growing demand for the STEM workforce worldwide. For example, the report by the U.S Bureau of Labour Statistics in 2017 stated that the demand for STEM-related workforce has increased to 10.5% from the period of May 2009 to May 2015 as compared to only 5.2% net growth in the non-STEM related workforce for the same period (Fayer et al., 2017). The same trend is projected in Malaysia as Malaysia is set to have one million qualified individuals to fill the STEM workforce demand by 2020, compared to the current

120,000 STEM workforce (Academy of Sciences Malaysia, 2015). Among the efforts undertaken by the Malaysia Ministry of Education (KPM) to increase skills and expertise in research and industry is through the strengthening of STEM education (Rasid, Nasir, Singh, & Cheong, 2020).

The success of this high demand for STEM careers in the future workforce is based on the country's current education system. This may be the reason why the Education Ministry of Malaysia had taken the initiative to include STEM in the curriculum in 2017, although it was done slightly later compared to other developed countries (Banks & Barlex, 2014; Curriculum Development Division, 2016a). Nonetheless, there are many things to consider in ensuring the success of STEM education such as that related to subject matter experts, resources, infrastructure, STEM teachers and most importantly the teachers' instructional preparedness. This article was written to discuss the development and validation of an instrument to measure STEM teachers' instructional preparedness.

2. The Evolution of STEM Education in Malaysia

While STEM-focused education had been introduced into the Malaysian curriculum since 2017, the effort to attract students into science related disciplines had started as far back as 1967, in a 60:40 ratio on Science-Art Policy (Ministry of Education, 2013). However, this policy failed due to many reasons, such as the ad-hoc changes in education policies, quality of teachers and ineffective STEM teaching methods (Academy of Sciences Malaysia, 2015). For example, a study conducted by Suhanna Zainudin et. al (2015) reported that students themselves state that the method of instruction used by science teachers is outdated and does not attract the students' interest in science and mathematics which are the main disciplines in STEM. This has significantly affected the enrolment rate of students in science courses in Malaysian schools.

Similarly, teachers were also reported lamenting on the difficulty in merging the various disciplines in STEM education as being one of the pedagogical problems in its teaching and learning. Teachers continue to teach STEM disciplines as silo subjects with little or no connection between one STEM subject and another. In addition, teachers also feel the lack of mastery as a subject matter experts of STEM subjects (Asghar et al., 2012; Kelley et al., 2016; Siew et al., 2015).

Another issue is the inappropriate selection and use of teaching aids, such as models, pictures, integration of information communication technology (ICT) and other visual representations by the teachers. Despite their significant effect on the success of STEM teaching and learning owing to the abstract nature of the concepts, these instructional aids are either absent or merely neglected in STEM classrooms (Ibrahim Abdulrahman Alkhaldi et al., 2020; Nur Farhana Ramli & Othman Talib, 2017).

Other challenges related to STEM education are the teachers' lack of expertise in assessment and evaluation. This is because for effective assessment of the STEM students' feedback, teachers need to use appropriate evaluation instruments (Asghar et al., 2012; Chandler et al., 2011). However, one of the biggest problems in curriculum development arises when the poor assessment method is applied and the assessment focuses only on the cognitive domain (Ejiwale, 2013).

In Malaysia, whilst the new STEM approach has been introduced and the evaluation standards have been proposed to the teachers (Curriculum Development Division, 2016b), proper training and professional development should be given regularly to the teachers since they are the first respondents to the students and need to be responsive and equipped in providing proper feedback to the students (Asghar et al., 2012; M. M. Capraro & Jones, 2013; Dodeen et al., 2012).

Teachers being core advocates in the curriculum implementation need to play an important role in ensuring that the implementation runs smoothly. Teachers who want to succeed in the classroom need to be adequately prepared. The studies prove that that instructional preparedness would have an impact on students achievement (Ku et al., 2020; Mohd Erfy et al., 2019; Slavit et al., 2016; Steele et al., 2012), Even though the study about STEM in Malaysia has been slowly increasing (Ahmad Zamri et al., 2017; Edy Hafizan et al., 2016; Fazilah et al., 2020; Nor Tutiaini & Othman, 2017), however, there are none on instruments measuring teachers' instructional preparedness. So, there is a need for the review and the development of instruments to measure STEM teachers' instructional preparedness.

2.1   Past Studies on Development of STEM Measurement Instruments

Previous studies in instrument development in STEM have been limited to college-level preparedness (Benjamin et al., 2015), self-efficacy (Mobley, 2015) and STEM discipline in silo research (Beerer & Bodzin, 2003; Ford, 2018; Yoon et al., 2014). Along with that, there are also some issues arising regarding the instructional development processes such as the scale used, cross culture translation and the instrument validation process (Hair et al., 2013; Nadelson et al., 2013; Zainudin Awang et al., 2016).

The present study is prompted with a scope to design, develop and validate an alternative instrument for the measurement of Malaysian science secondary school teachers' level of instructional preparedness in the implementation of STEM education. This is to ensure the STEM teachers in Malaysian Secondary schools are comprehensively evaluated for effective policy enactment for the successful implementation of STEM education.

3.   Methodology

3.1   Research Design

The instrument development and evaluation processes of an instrument in this study was designed by adapting the instrument development process proposed by DeVellis (2017), Nasab et al.(2015) and Miller et al. (2013). The process involves ten steps that are grouped into two phases namely, development and validation.
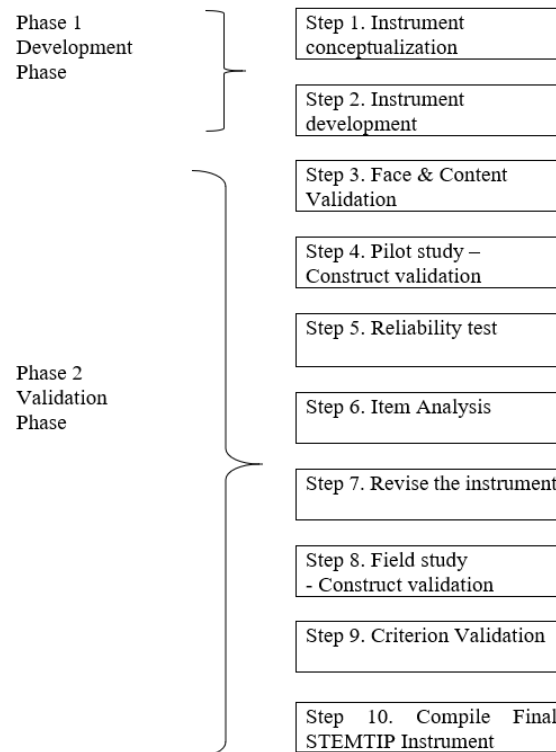

Fig. 1 Instrument development and validation process

Phase 1: Development Phase

Phase one involves the conceptualization and item generation process. Both of these were carried out through inductive and deductive processes. This means, apart from the analysis of literature review by researchers, the development of this instrument is also facilitated through focus group discussion. Four science secondary school teachers were selected as members of the focus group. For the mini group discussion, four participants are considered as sufficient (Greenbaum, 1993). All the documents related to STEM education in Malaysia from 2013 to 2017 were analyzed by researchers and were discussed in focus group discussion.

Based on the stated process, the concept of the development of instrument or tool was based on social constructivist theory (Belland, 2017), 5E instructional model (Bybee et al., 2006) and STEM teaching and learning approach model (Curriculum Development Division, 2016b). The 5E Instructional Model is the most suitable model for adaptation to all subjects in the national curriculum. The teaching materials module for STEM subjects is also using the 5E Instructional Model as a basis of the modules (Curriculum Development Division, 2017).

In line with that, the role of teachers as facilitators in STEM teaching and learning is very much emphasized in the guide. Besides teachers, peers as well as media assistance are considered important mediums to scaffold students to learn. This is consistent with the concept of scaffolding in social constructivist theory which promotes the role of these three parties in fulfilling students' zone of proximal development, where they need the right amount of guidance to complete the task successfully.

Based on the theoretical framework of this study, five constructs were developed as the core of this instrument. They are engagement, exploration, explanation, elaboration and evaluation. Fifty-one items were generated and were prepared for the validation process.
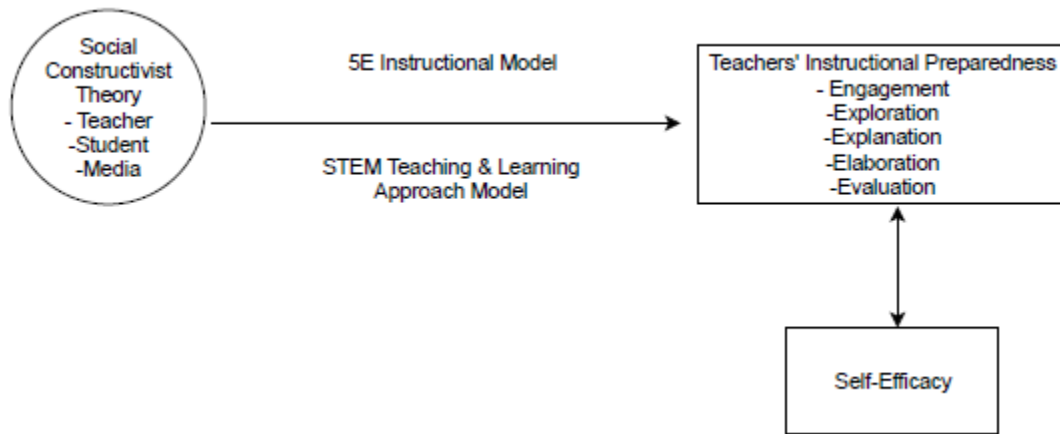

Fig. 2 Conceptual framework of STEMTIP

Apart from developing the item pools, the researcher also has planned and provided administration instruction for the respondents. This is to ensure the respondents can respond well and to get a high return rate. Both pilot and field study respondent selection was based on randomized multistage cluster sampling. The population of this study is science secondary school teachers as the science teacher is closest to the STEM teacher (El Nagdi et al., 2018). There were 45 science teachers from one state selected randomly for the pilot study. 489 science teachers from 56 schools from four states were selected for the field study. Mailed survey was used as a method in instrument distribution and respondents were asked to respond within the time frame given.

Phase 2: Validation Phase

The validation phase involves validity and reliability testing. Reliability refers to a measurement that yields consistent results every time it is being used (Miller et al., 2013; Zainudin Awang, 2015, Nasab et al.2015). As this instrument was analysed using the Rasch model, there were two additional reliability values obtained apart from Cronbach Alpha value. The reliability values are person and item reliability. Apart from giving information about the replicability of person and item placements along the trait continuum, these two values are also able to estimate the sufficiency of items and respondents used (Bond & Fox, 2015).

Validation is an important process to ensure that the developed instrument is able to measure what it intends to measure. Validity also refers to the ability to predict specific events, or its relationship to measure other constructs based on the manner in which a scale was constructed (DeVellis, 2017). The validity process includes face validity, content validity, criterion validity and construct validity (Bond & Fox, 2015; DeVellis, 2017).

Face validity is a process that requires selected respondents to evaluate the instrument based on the question interface, structure of sentences, grammar and other issues in the instrument that are deemed necessary. Despite testing the passable measures of the conceptual variables, it also helps the researcher to have early detection of the possibility of misunderstood or misinterpreted questions (Stangor, 2015; Zainudin Awang, 2015).

Content validation is the process of determining if the variables adequately cover the full measured domain, and this done by using the help of experts (Clark & Creswell, 2015; Stangor, 2015). In this process, the domain needs to be clearly defined in order to facilitate the evaluation process (DeVellis, 2017). There are no cut-off numbers in determining the number of experts, However, Zamanzadeh et al., (2015) suggest 5-10 content experts are recommended so as to have enough control over chance agreement.

In this study, construct validity was tested two times, during pilot and field study. Construct validity is an investigation to make sure the instrument developed measures correctly what it intended to measure. It predicts measured variables using several indicators, i.e. item fit or suitability of item, unidimensionality, local independence, item polarity and separation index. The Rasch model is used to measure construct validity. It was chosen for its ability to give more information apart from that from Classical Test Theory, such as the ability to discriminate items and person in one measurement scale, scale determination and quantitative item assessment (Bond & Fox, 2015; Cappelleri et al., 2014; Petrillo et al., 2015).

The four-point scale used in this instrument was "never", "rare", "some of the time", and "always". Respondents' choice was analyzed and converted to logit value to comply with rules in parametric tests which require interval data (Hair et al., 2013).

The final validation process involves criterion validity. Criterion validity occurs when the instrument has empirical association with some criterion or standard (DeVellis, 2017; Nasab et al.2015). There are two types of criterion validity, which are predictive and concurrent validity. Since there is no existing instrument in measuring instructional preparedness the use of predictive validity is advisable.

Predictive validity involves the attempt to forecast the future or draw an inference. The previous study conducted by Bruder et al. (2013) found that teachers' instructional preparedness is a predictor of teachers' self-efficacy. So, the newly developed STEMTIP instrument was tested to see its predictor effect to teachers' self-efficacy for predictive validity, using the Teachers' Self-Efficacy to Teach Science instrument (SETIS) (Mobley, 2015). As the instrument is used to test the predictive validity, it is necessary to develop regression analysis to confirm it (Gregory, 2004; Mohammad Rahim Kamaluddin & Rohany Nasir, 2017).

4. Results

4.1 Conceptualization of STEMTIP

Based on the literature review and the focus group discussion, Table 1 showed the conceptual and operational definition for the instrument for each of the constructs. All the items are previously written in Bahasa Melayu as it is easier to be understood by all experts and respondents.

Table 1. Conceptual and operational definition of STEMTIP

Construct (Conceptual Definition)	Aspect (Operational Definition)
Engagement Teachers' ability in preparing activities that promote curiosity, motivation and elicit students' prior knowledge with STEM activity.	Ability to attract student interest Ability to spark student curiosity Assess student prior knowledge Ability to associate student prior knowledge with lesson objective Ability to motivate student with STEM activity
Exploration Teachers' ability in designing STEM activities. Teachers also help students in providing a platform for information sources, probing questions, equipment and exploration materials and encourage them to collaborate.	Ability to search related information before the lesson Ability to plan STEM integrated exploration activity Ability to prepare basic material for the exploration activity Ability to scaffold student learning Ability to organize students to collaborate
Explanation Teachers' ability in defining new STEM concepts, process or skills and correcting misconception during mannered students' discussion. Teachers devote sufficient time and be fair to each student to share product / product summary and encourage students to make critical, critical, creative and fair judgments.	Ability to be a moderator in the discussion Ability to correct misconception Ability to use various types of platform for communication Ability to promote the integration of various information before making decisions Ability to define new STEM concept, process or skills with current situation
Elaboration Teachers' ability to challenge and extend students' conceptual understanding and skills through additional activities and questions that encourage students to think	Ability to prepare questions based on different level thinking skills Ability to use different types of questions techniques Ability to use gamification in learning Ability to encourage students to solve the problem Ability to accept possibilities to the answer Promoting community involvement in STEM problem solving
Evaluation Teachers' ability to evaluate student's progress throughout the teaching and learning session based on standard documents, assessing student's interest in STEM, preparing the report and discussing the progress with the suggestions on improvements.	Have standard evaluation instrument Being trained to evaluate students Ability to give improvement suggestions Ability to assess students' STEM interest Report the findings to the stakeholders

4.2 Participants

A total of 489 sets of questionnaires were distributed to the sample based on randomized multistage cluster sampling. However, only 265 teachers successfully returned the questionnaires. The Demographics profile in Table 2 recorded the highest respondent involvement in Kelantan with 85 teachers (32.1%), followed by 76 teachers from Melaka (28.7%), 67 teachers in Selangor (25.3%) and 37 teachers in Kedah (14.0%). Thirty-five (13.2%) respondents were male teachers and 230 (86.8%) were female teachers. The majority of the respondents were teachers in 134 (50.6%) science subjects, followed by 49 biology teachers (18.5%), 41 physics teachers (15.5%) and 41 chemistry teachers (15.5%).

Table 2. Demographics profile of the respondents

	Frequencies	**Percentage (%)**
State		
Kelantan	85	32.1
Melaka	76	32.1
Selangor	67	25.3
Kedah	37	14.0
Gender		
Male	35	13.2
Female	230	86.8
Subject teach		
Science	134	50.6
Biology	49	18.5
Chemistry	41	15.5
Physic	41	14.5
Total	**265**	**100**

4.3 Reliability

Reliability refers to the repeatability of a measure. Three reliability values obtained using the Rasch model were summarised in Table 3.

Table 3. STEMTIP reliability indices

Type of reliability	Value
Cronbach Alpha (KR-20)	.98
Person reliability	.96
Item reliability	.98

4.4 Content validity

Content validity concerns item sampling adequacy to test whether a specific set of items reflects the domain content. There were 51 items which were initially generated from the inductive and deductive processes. The items then went through the face validation and content validation process. Five science teachers were appointed to examine the sentence structure, spelling and suitability in terms of the use of words for each item in face validation. The items were then assessed by the group of professionals as well as lay experts for content validity. Seven professional experts were appointed from various backgrounds such as in the STEM, language and psychometric areas. Eight science teachers were selected as lay experts as they were specialists in STEM implementation in school.

Each of the items was rated based on the Lawshe Model (Lawshe, 1975; Wilson et al., 2012). The Lawshe Model used a three-point scale in reviewing each individual item; (1) essential, (2) useful but not essential and (3) not necessary. Then, the responses from all the experts were pooled and the number indicating "essential" for each item was determined. The critical value of content validity ratio was determined using formula equation (1)

$$CVR = [n_e - (N / 2)] / (N / 2) \text{ -------------------(1)}$$

Where, n_e is the number of experts indicating "essential" and N is the total number of experts.

Based on the consensus from the experts, 47 items were accepted to reflect the concept of STEM teachers' instructional preparedness. However the remaining four items were carried forward to the next pilot and field study based on the experts' recommendations and focus group discussion endorsement after a few modifications (Nur Farhana Ramli et al., 2018).

4.5   Construct validity

Construct validity is performed to ensure the instrument measures what is expected to measure. For this study, the data were analysed by the Rasch model. There are five assumptions in the Rasch Model in verification of construct validity. The assumptions are item fit, unidimensionality, local independence, item polarity and separation index. The results are summarised in Table 4.

Table 4. Construct Validity Results

Rasch model assumptions	Purpose		Results	Quality	References
Item fit	Item fit testing was conducted to make sure the item are fit	Infit MNSQ	0.78 to 1.25	Excellent	(Fisher, 2007)
	in the model, which means it gives information to the measurement	Outfit MNSQ	0.78 to 1.28	Excellent	(Fisher, 2007)
Unidimensionality	Unidimensionality test was conducted to make sure the instrument is measuring only one dimension	Principal components analysis (PCA)	46.1%	Exceed min 40%	(Linacre, 2018) (Azrilah et al., 2017)
		Noise Variance ratio	4.9% 9.4:1	Very good Unidimensionality	(Fisher, 2007) (Linacre, 2018)
		Eigen	3.7	Signal multidimensionality	(Linacre, 2018)
Local independence	Local independence test was conducted to make sure that the items are not related to each other.		.36 to .57	Locally dependent	(Linacre, 2018)

Item polarity	Item polarity is an indicator used to indicate the measured items will move in one direction in the constructed dimension.		0.51-0.67	Unidimensionality	(Linacre, 2018) (Bond & Fox, 2015)
Separation index	Separation index provides information about the ability of the instrument to discriminate items and persons.	Person	6.63	Excellent	(Fisher, 2007)
		Item	9.04	Excellent	(Fisher, 2007)

4.6   Criterion validity

Criterion validity occurs when the instrument has an empirical association with some criterion or standard. In this study, STEMTIP was tested to see whether it has empirical association with STEM teachers' self-efficacy (SETIS). Data was analysed using simple linear regression, and the results are shown in Table 5.

In this analysis, STEMTIP was the predictor of STEM teachers' self-efficacy. Based on the adjusted R square value, it showed that STEMTIP was a predictor to 40.6% of STEM teachers' self-efficacy.

Table 5. Regression analysis summary

Model	R	R square	Adjusted R square	Standard error
1	.639	.409	.406	2.03783

Note: Dependant variable: SETIS, Predictors (Constant): STEMTIP

5.   Discussion

STEMTIP was developed for the purpose of measuring teachers' instructional preparedness in STEM implementation. Apart from its importance in determining teachers' preparedness, this measurement is also a necessity to ensure the success of the national education aspirations. This can indirectly meet Malaysia's STEM future workforce demand.

There were five constructs being developed as the core of the instrument. These are engagement, exploration, explanation, elaboration and evaluation. Fifty-one items were initially generated and went through the face and content validation processes. Even though there were four items that failed to get the consensus from the group of experts, there were still included in the pilot and field study after a few modifications and suggestions resulting from the focus group discussions.

Based on the reliability value that was above 0f 0.96, STEMTIP can be considered as an excellent instrument (Fisher, 2007). It is proof that this instrument is able to be replicated in another sample of science teachers in Malaysia. It also indicates that sufficient items and samples are being used in this study (Linacre, 2012).

To ensure that the items generated measure the dimensions of teacher instructional preparedness in STEM implementation, construct validation was executed. The Rasch model was used as the basis for this measurement. There are five assumptions that need to be met to ensure construct validity. The assumptions are item fit, unidimensionality, local independence, and item polarity and separation index.

The results of the Rasch analysis demonstrate that the STEMTIP has good psychometric qualities. In the determination of fit items, misfit items need to be re-considered because of their ability to interfere with the measurements. Based on the test, ten items were identified as misfit items and were deleted where they were over the range of 0.77 to 1.3 logit that was suggested by (Fisher, 2007). The ten items were summarised in Table 6.

Table 6. Deleted misfit items

No	Item
1	I integrate STEM in activities to attract student engagement
11	I am integrating Science, Technology, Design and Mathematics in exploration activities
24	I can guide the discussion of STEM teaching and learning activities
39	I accept various answers that were included with justification.
41	I make sure students work together in problem solving activities
25	I master all the titles contained in the DSKP (standard curriculum and assessment documents) of the subject being taught
37	I use gamification activity (example: educational video games, Kahoot, and Quizizz) to reinforce student understanding
42	I involve stakeholders (parents, community members, and field experts) in problem solving activities
45	I have the ability to assess STEM teaching and learning activity
51	I discussed the STEM students' achievement to the stakeholders (eg parents, schools, MOE)

The remaining 41 items were of excellent qualities in measuring teachers' instructional preparedness. This value is determined based on the logit value obtained from the respondents. The list of final STEMTIP items can be accessed here http://bit.ly/STEMTIP.

The Rasch model, apart from proving that the instrument is measuring one dimension, also emphasizes that there is no existence of a second dimension. There are four assumptions to be met in the determination of unidimensionality. Even though there was a signal on the second dimension based on the eigenvalue (see Table 4), the other three assumptions, PCA value, noise value, and variance ratio are enough to accept the instrument unidimensionality (Azrilah Abdul Aziz et al., 2017; Fisher, 2007; Linacre, 2018).

The next test for teachers' instructional preparedness is local independence. Local independence is the assumption that the response to one item should not lead to a response to another item. Based on the results, all the items in STEMTIP was proven to be locally independent. The highest correlation value between the items is .57 which is below the cut-off point suggested by Linacre (2018). All positive value in item polarity is another indicator of unidimensionality which shows that all the items are measuring one latent construct (Bond & Fox, 2015; Linacre, 2012).

The last indicator is the separation index test. A good set of test items can be discriminated against by respondents (Cohen et al., 2013). The Rasch model is able to indicate the discrimination through person separation index and item separation index. Person separation index aims to estimate the ability of the instrument to separate the person to several groups in the measured constructs, while item separation index aims to estimate a person's ability by separating item difficulty to few groups in the measured constructs. STEMTIP is able to discriminate between six groups and items to nine groups. Based on the compliance of all the Rasch model assumptions, it can be said that STEMTIP had a good indicator of construct validity.

The last validity test is criterion validity. In this test, STEMTIP proved to predict STEM teachers' self-efficacy through teachers' instructional preparedness. Based on the regression analysis that was conducted, STEMTIP was proven to predict 40.6% of STEM teachers' self-efficacy (see table 5). This

value exceeds the 30% value which is sufficient value for criterion validity (Nunnally & Bernstein, 1994). This undeniably proves STEMTIP criterion validity.

6.   Implication and Conclusion

Although teacher's preparedness is important in teaching and learning there is a lack of empirical studies on STEM teachers which focus especially on instructional preparedness. This article discussed the process of development and validation of the newly developed STEM Teachers' Instructional Preparedness (STEMTIP) instrument. In the process of developing this instrument, the conceptualization process is crucial to ensure that all items constructed can measure a latent variable. From the concepts constructed, items were generated based on the operational definition. In this study, 41 out of 51 items have been verified to measure STEM teachers' instructional preparedness through reliability testing, content validity, construct validity and criterion validity.

Further contributions and its implications are as follows. Firstly, based on the excellent reliability values which range from .96 to .98, it can be concluded that the instrument can be replicated to another set of teachers in Malaysia. The sampling method used, i.e. randomization multistage cluster sampling reinforces that the invariance sample can be obtained through this method. Even though only 265 teachers (54.2%) have successfully returned the questionnaires to be analysed, the number of respondents is adequate based on Linacre's (1994) suggestion. The invariance sample is very important in producing a good instrument, which can discriminate persons in it.

Secondly, information from items discrimination found that the most difficult item to endorse for teachers was the item that relates to engineering disciplines in STEM exploration activities. This information was useful information for stakeholders in relevant training planning in the future. Surely, without content knowledge, teachers cannot perform effective teaching and learning sessions. The stakeholders also recommended the need to conduct an in-depth study to find out the problem of teachers towards the engineering discipline.

In addition, based on the predictive validity, test results of the instrument showed that there is a relationship between teachers' instructional preparedness and teachers' self-efficacy. Self-efficacy is a personal belief in his or her ability to succeed in a situation. That means, the teacher needs to be prepared to increase teacher's self-efficacy in teaching STEM.

Lastly, the use of the Rasch model in construct validity testing proves to give more information from other methods. Items have been tested individually to conform the assumptions in the model through item fit test, unidimensionality, local independence, item polarity and separation index. Item and person reliability are also an added value that can prove the adequacy and variance of the sample.

Overall, the development of this instrument has succeeded in producing an instrument that can measure teachers' instructional preparedness in STEM implementation through the rigorous steps on the development and validation processes. The findings of this instrument are expected to add more knowledge in the STEM field as well as in teacher instructional preparedness, especially in Malaysia context.

7.   References

Academy of Sciences Malaysia. (2015). *Science Outlook 2015: Action Towards Vision* (Academy of Sciences Malaysia (ed.)). Ministry of Science, Technology And Innovation.

Ahmad Zamri Khairani, Khairani, A. Z., & Ahmad Zamri Khairani. (2017). Assessing Urban and Rural Teachers' Competencies in STEM Integrated Education in Malaysia. *MATEC Web of Conferences*, *87*, 04004. https://doi.org/10.1051/matecconf/20178704004

Asghar, A., Ellington, R., & Rice, E. (2012). Supporting STEM education in secondary science contexts. *Interdisciplinary Journal of Problem- Based Learning*, *6*(2), 85–125. https://doi.org/10.7771/1541-5015.1349

Azrilah Abdul Aziz, Mohd Saidfudin Masodi, & Azami Zaharim. (2017). *Asas Model Pengukuran Rasch : Pembentukan Skala dan Struktur Pengukuran* (1st ed). Penerbit Universiti Kebangsaan Malaysia.

Banks, F., & Barlex, D. (2014). Teaching STEM in the Secondary School. In *Teaching STEM in the Secondary School: Helping Teachers Meet the Challenge*. https://doi.org/10.4324/9780203809921

Beerer, K., & Bodzin, A. (2003). *Science Teacher Inquiry Rubric (STIR)*. http://www.lehigh.edu/~amb4/stir/stir.pdf.

Belland, B. R. (2017). *Instructional scaffolding in STEM education : strategies and efficacy evidence*. Springer. https://doi.org/10.1007/978-3-319-02565-0

Benjamin, T. E., Marks, B., Demetrikopoulos, M. K., Rose, J., Pollard, E., Thomas, A., & Muldrow, L. L. (2015). Development and Validation of Scientific Literacy Scale for College Preparedness in STEM with Freshman from Diverse Institutions. *International Journal of Science and Mathematics Education*, 1–17. https://doi.org/10.1007/s10763-015-9710-x

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model : fundamental measurement in the human sciences* (3rd ed., Vol. 1, Issue 3). Routledge. https://doi.org/10.1207/S15327574IJT013&4_10

Bruder, M. B., Dunst, C. J., Wilson, C., & Stayton, V. (2013). Predictors of confidence and competence among early childhood interventionists. *Journal of Early Childhood Teacher Education*, *34*(3), 249–267. https://doi.org/10.1080/10901027.2013.816806

Bybee, R. W., Taylor, J. A., Van Scotter, P., Powell, J. C., & Landes, N. (2006). *The BSCS 5E Instructional Model : Origin and Effectiveness*.

Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, *36*(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006

Capraro, M. M., & Jones, M. (2013). Interdisciplinary STEM project-based learning. In R. M. Capraro & M. M. Capraro (Eds.), *STEM project-based learning : An integrated Science, Technology, Engineering and Mathematics (STEM) approach* (2nd ed., pp. 47–54). Sense Publishers.

Chandler, J., Fontenot, a. D., & Tate, D. (2011). Problems Associated with a Lack of Cohesive Policy in K-12 Pre-college Engineering. *Journal of Pre-College Engineering Education Research (J-PEER)*, *1*(1), 5. https://doi.org/10.7771/2157-9288.1029

Clark, V. L. P., & Creswell, J. W. (2015). *Understanding Research : A Consumer 's Guide* (2nd ed.). Pearson Education.

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment: an introduction to tests and measurement* (8th ed.). McGraw Hill.

Curriculum Development Division. (2016a). *Buku Penerangan Kurikulum Standard Sekolah Menengah (KSSM)*. Ministry of Education.

Curriculum Development Division. (2016b). *Panduan pelaksanaan Sains, Teknologi, Kejuruteraan dan Matematik (STEM) dalam pengajaran dan pembelajaran*. Kementerian Pendidikan Malaysia.

Curriculum Development Division. (2017). *Bahan sumber STEM*. Ministry of Education. http://bpk.moe.gov.my/index.php/inisiatif-pppm/stem/buku-sumber-stem

DeVellis, R. . (2017). *Scale development : theory and application* (4th ed.). SAGE Publications Inc.

Dodeen, H., Abdelfattah, F., Shumrani, S., & Hilal, M. A. (2012). The Effects of Teachers' Qualifications, Practices, and Perceptions on Student Achievement in TIMSS Mathematics: A Comparison of Two Countries. *International Journal of Testing*, *12*(1), 61–77. https://doi.org/10.1080/15305058.2011.621568

Edy Hafizan Mohd Shahali, Ihsan Ismail, & Lilia Halim. (2016). STEM education in Malaysia: policy, trajectories and initiatives. *Asian Research Policy*, *8*(2), 122–133.

Ejiwale, J. A. (2013). Barriers to Successful Implementation of STEM Education. *Journal of Education and Learning*, *7*(2), 63–74.

El Nagdi, M., Leammukda, F., & Roehrig, G. (2018). Developing identities of STEM teachers at emerging STEM schools. *International Journal of STEM Education*, *5*(1), 1–13. https://doi.org/10.1186/s40594-018-0136-1

Fayer, S., Lacey, A., & Watson, A. (2017). *STEM Occupations : Past , Present , And Future* (Issue January).

Fazilah Razali, Umi Kalthom Abdul Manaf, & Ayub, A. F. M. (2020). STEM Education in Malaysia towards Developing a Human Capital through Motivating Science Subject. *International Journal of*

Learning, Teaching and Educational Research, *19*(5), 411–422. https://doi.org/10.26803/ijlter.19.5.25

Fisher, W. P. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions*, *21*(1), 1095. https://doi.org/10.1017/CBO9781107415324.004

Ford, C. (2018). Effective practice instructional strategies: Design of an instrument to assess teachers' perception of implementation. *Studies in Educational Evaluation*, *56*, 154–163. https://doi.org/10.1016/j.stueduc.2017.12.004

Greenbaum, T. L. (1993). *The handbook for focus group research* (2nd ed.). Lexington Books.

Gregory, R. J. (2004). *Psychological testing : history, principles and application* (4th rd.). Pearson.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). Multivariate Data Analysis. In *Vectors* (7th ed.). Pearson Education Ltd. https://doi.org/10.1016/j.ijpharm.2011.02.019

Ibrahim Abdulrahman Alkhaldi, Othman Talib, Habibah Ab Jalil, & Nurzatulshima Kamarudin. (2020). The Use of Mobile Application : Why Play is Important in Teaching Elementary Mathematics ? *Talent Development & Excellence*, *12*(3), 2945–2956.

Kelley, T. R., Knowles, J. G., Abell, S., Lederman, N., Bandura, A., Bandura, A., Barak, M., Barnett, J., Hodson, D., Breiner, J., Harkness, M., Johnson, C., Koehler, C., Berlin, D., White, A., Brown, J., Collins, A., Duguid, P., Burghardt, M., … Williams, D. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1), 11. https://doi.org/10.1186/s40594-016-0046-z

Ku Sook Leng, Fazilah Razali, & Ayub, A. F. M. (2020). The effectiveness of realistic mathematics education approach toward students learning: A systematic literature review of empirical evidence. *Journal of Critical Reviews*, *7*(13), 548–552. https://doi.org/10.31838/jcr.07.13.99

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/J.1744-6570.1975.TB01393.X

Linacre, J. M. (1994). Sample Size and Item Calibration Stability. *Rasch Measurement Transactions*, *7*(4), 328.

Linacre, J. M. (2012). *A user's guide to WINSTEPS: Rasch Model Computer Programs*. MESA Press. https://doi.org/ISBN 0-941938-03-4

Miller, L. A., Lovler, R. L., & McIntire, S. A. (2013). *Foundations of Psychological Testing : A Practice Approach* (4th ed.). SAGE Publications Inc.

Ministry of Education. (2013). Preliminary Report : Malaysia Education Blueprint 2013 - 2025. In *Kementerian Pendidikan Malaysia (MOE)*. https://doi.org/10.1016/j.tate.2010.08.007

Mobley, M. C. (2015). *Development of the SETIS Instrument to Measure Teachers' Self-Efficacy to Teach Science in an Integrated STEM Framework*. University of Tennessee.

Mohammad Rahim Kamaluddin, & Rohany Nasir. (2017). Kesahan kandungan dan muka. In Mohammad Rahim Kamaluddin & Wan Shahrazad Wan Sulaiman (Eds.), *Teknik kesahan dan kebolehpercayaan alat uji psikologi* (pp. 39–50). Penerbit Universiti Kebangsaan Malaysia.

Mohd Erfy Ismail, Suhaizal Hashim, Norhasyimah Hamzah, Noorazman Abd Samad, Saiful Hadi Masran, Khairul Azhar Mat Daud, Nor Fadila Mohd Amin, Mohd Ali Shamsudin, & Nur Zatul Shima Kamarudin. (2019). Factors that influence students' learning: An observation on vocational college students. *Journal of Technical Education and Training*, *11*(1), 93–99. https://doi.org/10.30880/jtet.2019.11.01.12

Nadelson, L. S., Callahan, J., Pyke, P., Hay, A., Dance, M., & Pfiester, J. (2013). Teacher STEM perception and preparation: inquiry-based STEM professional development for elementary teachers. *The Journal of Educational Research*, *106*(2), 157–168. https://doi.org/10.1080/00220671.2012.667014

Nasab, M.F., Abdul Kadir, R., Hassan, S.A. & Mohd. Noah, S. (2015). Psychometric Evaluation of the Career Decision Scale With Iranian Undergraduate Students. *Journal of Counseling & Development*. 93(3), 344 – 351

Nor Tutiaini Ab Wahid, & Othman Talib. (2017). STEM Integration in Classroom Practices among Biology Teachers in Mara Junior Science College (MJSC). *International Journal of Academic Research in Business and Social Sciences*, *7*(4), 1030–1041. https://doi.org/10.6007/IJARBSS/v7-i4/2912

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw Hill.

Nur Farhana Ramli, & Othman Talib. (2017). Can educational institutions implement STEM? From Malaysian teachers ' view. *International Journal of Academic Research in Business and Social Sciences*, *7*(3), 721–732. https://doi.org/10.6007/IJARBSS/v7-i3/2772

Nur Farhana Ramli, Othman Talib, Siti Aishah Hassan, & Umi Kalthom Abdul Manaf. (2018). Content Validity of STEMTIP Using CVR Method. *International Journal of Academic Research in Business and Social Sciences*, *8*(7), 1118–1125. https://doi.org/10.6007/IJARBSS/v8-i7/4559

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, *18*(1), 25–34. https://doi.org/10.1016/j.jval.2014.10.005

Rasid, N. S., Nasir, N. A., Singh, P. S., & Cheong, T. H. (2020). STEM Integration: Factors Affecting Effective Instructional Practices in Teaching Mathematics. *Asian Journal of University Education, 16*(1), 56. doi:10.24191/ajue.v16i1.8984

Siew, N. M., Nazri Amir, & Chong, C. L. (2015). The perceptions of pre-service and in-service teachers regarding a project-based STEM approach to teaching science. *SpringerPlus*, *4*(1), 1–20. https://doi.org/10.1186/2193-1801-4-8

Slavit, D., Nelson, T. H., & Lesseig, K. (2016). The Teachers' Role in Developing, Opening, and Nurturing an Inclusive STEM-focused School. *International Journal of STEM Education*, *3*(1), 7. https://doi.org/10.1186/s40594-016-0040-5

Stangor, C. (2015). *Research methods for the behavioral sciences* (5th ed.). Cencage Learning.

Steele, A., Brew, C. R., & Beatty, B. R. (2012). The tower builders: A consideration of STEM, STSE and ethics in science education. *Australian Journal of Teacher Education*, *37*(10), 118–133. https://doi.org/10.14221/ajte.2012v37n10.2

Suhanna Zainudin, Lilia Halim, & Zanaton Ikhsan. (2015). How 60 : 40 Policy Affects The Development Of Science Curriculum in Malaysia. *Proceeding: 7th International Seminar on Regional Education, November 5-7, 2015*, *3*, 1396–1405.

U.S. Information Agency. (1959). *Impact of US and Soviet space programs on world opinion* (Issue July).

Wilson, F. R., Wei Pan, & Schumsky, D. A. (2012). Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210. https://doi.org/10.1177/0748175612440286

Yoon Yoon, S., Evans, M. G., & Strobel, J. (2014). Validation of the teaching engineering self-efficacy scale for K-12 teachers: A structural equation modeling approach. *Journal of Engineering Education*, *103*(3), 463–485. https://doi.org/10.1002/jee.20049

Zainudin Awang. (2015). *Postgraduate research proposal: a step by step guide in writing proposals for postgraduate students* (1st ed). MPWS Rich Publication Sdn Bhd.

Zainudin Awang, Asyraf Afthanorhan, & Mustafa Mamat. (2016). The Likert scale analysis using parametric based Structural Equation Modeling (SEM). *Computational Methods in Social Sciences*, *4*(1), 13–21.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017