



## Measuring the academic success of students with ASICS using polytomous item response theory



Hishamuddin Ahmad \*, Siti Eshah Mokshein, Mohd Razimi Husin, Siti Rahaimah Ali, Ismail Yusuf Panessai

Department of Educational Studies, Faculty of Human Development, Universiti Pendidikan Sultan Idris, Perak, Malaysia

### ARTICLE INFO

#### Article history:

Received 5 December 2018

Received in revised form

15 February 2019

Accepted 15 February 2019

#### Keywords:

Polytomous item response theory

Fully residential school

Academic success

### ABSTRACT

Fully Residential School (FRS) under the Ministry of Education in Malaysia is a type of school, which produces high performance human capital. Important factors in the academic success of high performing students should be focused on to produce more high-performing students to fulfill the needs of the nation. This study aims to apply the Academic Success Inventory for College Students (ASICS) in the context of FRS in Malaysia and measure the instrument quality as well as identify the most important factor in the academic success. The sample comprised 305 students from three FRS. The data was collected using the ASICS instrument, which contained 49 items. Later, the data was analyzed based on the Polytomous Item Response Theory using the Xcalibre software. Based on the Chi-square, p-value, and the -2 LogLikelihood, Samejima's Graded Rating Model was found to be the fit model with the data. Unidimensionality assumption and local independence were tested using the exploratory factor analysis and were fulfilled. The instrument's reliability was overall very satisfactory ( $\alpha=0.89$ ) and the construct validity was also fulfilled with the value of 0.86. It was found that most items (91.8%) showed good discrimination. The findings also showed that (i) ASICS is a good instrument to measure academic success among the students as previous studies and (ii) high performing students found to prioritize in aspects such 'Perceiving Instructor Efficacy' and 'Personal Adjustment' to achieve their academic success. It means, to develop an excellent future generation, we require an efficient education sector, especially one which focuses on efficient teachers for the 4.0 Industrial Age.

© 2019 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Sekolah Berasrama Penuh (SBP) or known as Fully Residential School (FRS) is a type of school which takes in students who have done extremely well in primary school examinations (UPSR or *Penilaian*) or secondary school examinations (PT3, PMR, or SRP). The students which have been included into the FRS's learning system can be considered as those who are excellent in academic and co-curriculum aspects. According to Ghani et al. (2013), FRS produces students who have a 'towering personality', a term defined as having excellent skills, knowledge, and moral values. In other words, FRS is a place where high performing students with excellent academic achievement are grouped

together at secondary education level and they are expected to achieve excellent results in *Pentaksiran Tingkatan 3* (PT3) and *Sijil Pelajaran Malaysia* (SPM).

The institution that takes in the students through a selection process will apply a variety of methods to filter the candidates. The methods used in these schools include analyzing previous results of centralized examinations, conducting interviews, and providing response for one or more instruments, which measure the desired characteristics. Typically, the selected students with their own characteristics to further drive their academic abilities will also highlight the institution's excellence.

There are a few requirements for measuring the overall and comprehensive academic achievement to assess many academic factors. An instrument known as Academic Success Inventory for College Students or ASICS developed by Prevatt et al. (2011) is utilized to measure factors related to academic success. To measure the quality of the instrument and review the factors, which have contributed towards the students' success, an analysis based on

\* Corresponding Author.

Email Address: [hishamuddin.a@fpm.upsi.edu.my](mailto:hishamuddin.a@fpm.upsi.edu.my) (H. Ahmad)

<https://doi.org/10.21833/ijaas.2019.04.014>

Corresponding author's ORCID profile:

<https://orcid.org/0000-0002-7120-1593>

2313-626X/© 2019 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

the Polytomous Item Response Theory (Polytomous IRT) by [Ostini and Nering \(2006\)](#) was utilized.

## 2. Objectives

This study aims to apply the ASICS instrument in the context of FRS in Malaysia and also to measure the instrument quality and identify important factors in the academic achievement of the high-performing students. Briefly, the study objectives are listed below:

- To assess the reliability and construct validity of instrument.
- To identify the item discrimination parameter.
- To assess the theta score pattern towards the constructs.

## 3. Methodology

### 3.1. Population and sampling

As shown in [Table 1](#), there are 69 fully residential secondary schools in Malaysia ([KPM, 2017a](#)). According to the [KPM \(2017b\)](#), the enrolment of students from Form 1 to Form 3 (lower secondary) in the whole of Malaysia was 23,238 on 30 June 2017. On the other hand, the enrolment of students from Form 4 and Form 5 was 15,021. Overall, the enrolment of students at secondary level (Form 1 to Form 5) was 38,259 ([KPM, 2017b](#)).

**Table 1:** Fully residential schools (Government – aided) in Malaysia

Number	State	Number of FRS
1	Perlis	1
2	Kedah	5
3	Pulau Pinang	2
4	Perak	7
5	Wilayah Persekutuan Putrajaya	1
6	Wilayah Persekutuan Kuala Lumpur	3
7	Selangor	9
8	Negeri Sembilan	8
9	Melaka	2
10	Johor	7
11	Pahang	7
12	Terengganu	6
13	Kelantan	5
14	Sabah	2
15	Wilayah Persekutuan Labuan	1
16	Sarawak	3
	Total	69

Negeri Sembilan has 125 secondary schools under the KPM ([KPM, 2018](#)). According to the [KPM \(2017b\)](#) besides FRS, these schools include Special Model School, Sports School, Guidance School, Art School, Religious Schools, Government funded Religious School, Special Education School, and Vocational College. The statistics compiled by the [KPM \(2017b\)](#) showed the enrolment of 83,237 students in 125 secondary schools from Form 1 to Form 5 on 30 June 2017 in Negeri Sembilan. [KPM \(2018\)](#) stated that out of the 125 secondary schools, 8 of them are in the FRS category ([Table 2](#)). Starting in 2014, the *Peperiksaan Menengah Rendah* (PMR)

examinations in Form 3 was replaced with *Pentaksiran Tingkatan 3* (PT3). In line with the implementation, the school ranking based on PT3 results was not announced publicly like what was done in previous years. As such, to assess the achievement of a school, this study would be focusing on the school rankings based on the SPM results.

Based on the SPM results for year 2015 and 2016 ([Table 2](#)), Negeri Sembilan has three FRSs which were in the top 10 ranking from all FRSs in Malaysia: Kolej Tunku Kurshiah, Sekolah Menengah Sains Tuanku Munawir, and Sekolah Menengah Sains Perempuan Seremban. These three schools are located in the Seremban district, Negeri Sembilan. This shows that these schools share an environment which is conducive for moulding the students towards excellent achievement. Additionally, there is another FRS in Seremban, Sekolah Dato' Abdul Razak which also has a good ranking in year 2015 SPM results.

**Table 2:** The ranking of FRS in Negeri Sembilan based on SPM results

No.	FRS in Negeri Sembilan	FRS Rank	
		SPM 2015	SPM 2016
1.	SM Agama Persekutuan Labu, Labu	37	15
2.	Kolej Tunku Kurshiah, Seremban	10	1
3.	Sekolah Dato' Abdul Razak, Seremban	27	40
4.	SM Sains Tuanku Jaafar, Kuala Pilah	41	36
5.	SM Sains Tuanku Munawir, Seremban	1	5
6.	SBP Integrasi Jempol, Jempol	34	42
7.	SM Sains Rembau, Rembau	33	32
8.	SM Sains Perempuan Seremban	5	8

In Negeri Sembilan, the enrolment of students into the fully residential schools was 4724 students with 2798 at the lower secondary level and 1926 at the higher secondary level. This represents 12.34 percent (4724 from 38,259) of the overall enrolment of students in FRS. This can be considered a high percentage as Negeri Sembilan has eight out of 69 FRSs (11.59%) in the whole of country.

As this research took into account the data collection which was conducted at the beginning of the year, it was unsuitable to involve Form 1 students because they had just entered the FRS system. This study also did not take into account students who are in the examination classes in current year (Form 3 and Form 5) as stated by [MOE \(2018\)](#). Therefore, this study is focused on Form 2 students only. During the data collection phase, the researchers sought the permission from four FRSs in Seremban: Kolej Tunku Kurshiah, Sekolah Menengah Sains Tuanku Munawir, Sekolah Menengah Sains Perempuan Seremban, and Sekolah Dato' Abdul Razak. However, Sekolah Menengah Sains Perempuan Seremban later decided to withdraw from the study. The three schools which had agreed to become the respondents were not only categorized as FRS but also *High Performing Schools* (HPS).

As stated by [KPM \(2018\)](#), HPS is defined as a school which has a unique ethos, character, and

identity in all education aspects. Such a school has a high achieving work culture which enables the development of the nation's human capital holistically and sustainably, as well as being competitive in the international arena. These made HPS as the chosen school among the Malaysian public.

The population of Form 2 students in Kolej Tunku Kurshiah, Sekolah Menengah Sains Tuanku Munawir, and Sekolah Dato' Abdul Razak can be seen in Table 3.

**Table 3:** Population and study sample from FRS in Seremban district

No.	FRS in Seremban	FRS Rank	
		Population	Sample
1.	Kolej Tunku Kurshiah	143	105
2.	SM Sains Tuanku Munawir	119	109
3.	Sekolah Dato' Abdul Razak	127	91
	Total	389	305

Due to the constraints related to learning activities at school, the researchers only managed to get a less number of respondents. The sample acquired was 305 respondents from the population of 389 Form 2 students from the three schools. According to Krejcie and Morgan (1970), the sample required for a population of 389 students at 95 percent confidence level is 194 students.

However, in the IRT context, a sample which closely resembles the actual population in terms of numbers is preferred to describe the findings of the study. Furthermore, according to Rivers et al. (2009), by assessing the relationship between latent trait and item properties, IRT effectively controls the differences in latent trait. Therefore, no random sampling is required

DeMars (2010) found that a sample size of 300 was required for an item calibration with a polytomous IRT model. In fact, if the sample size was small or less than 300, Guyer and Thompson (2013) explained that the *chi-square* ( $\chi^2$ ) fit statistics used in a polytomous IRT model would always provide statistically insignificant *p* values. If such a thing happens, it will certainly provide a meaningless interpretation to the analysis results. Therefore, a

sample of 305 is considered adequate to make a generalization of the population in this study.

### 3.2. Instrument

ASICS is an instrument copyrighted 2011 by Dr. Frances Prevatt with originally 50 items consisting 10 factors: General Academic Skills, Career Decidedness, Internal Motivation/Confidence, External Motivation/Future, Lack of Anxiety, Concentration, Socializing, Personal Adjustment, Perceived Instructor Efficacy, and External Motivation/Current.

Study conducted by Prevatt et al. (2011) found that the factors or construct of General Academic Skills had the highest internal consistency, while the External Motivation/Current showed the lowest internal consistency. They also found that Personal Adjustment, General Academic Skills, Internal Motivation/Confidence, Socializing, and Concentration were the most highly predictive subscales of grade point average (GPA).

To conduct this study, permission was granted to use the measure on May 20, 2014. Although Cohen and Swerdlik (1992) suggested that the construction of instruments involve phases such as planning, construction, testing, and validation, the instruments used in this study are adapted with permission without involving planning and construction phases. However, the expert's confirmation of the items has been obtained to fit the 'climate' of Malaysia.

According to the expert who had evaluated the instrument, one item related to 'drink', which referred to an alcoholic drink had been excluded. This was necessary as the expert viewed that it was not suitable with the culture and environment of the FRS students who were all under 18 years old and of the Muslim faith. As such, the remaining items were 49 items. For each item, the respondents were required to provide responses based on the Likert scale from '1' (strongly disagree) to '7' (strongly agree). The information pertinent to constructs and its items are shown in Table 4.

**Table 4:** Construct and study items

Code	Construct	Number of items	Item
1	Career Decidedness	4	Item46, Item47, Item48, Item49
2	Internal Motivation / Confidence	8	Item6, Item9, Item10, Item11, Item18, Item20, Item29, Item30
3	External Motivation / Future	4	Item7, Item19, Item38, Item41
4	General Academic Skills	12	Item4, Item8, Item12, Item14, Item23, Item31, Item33, Item34, Item42, Item43, Item44, Item45
5	Lack of Anxiety	3	Item3, Item15, Item32
6	Concentration	4	Item2, Item5, Item16, Item21
7	External Motivation / Current	3	Item26, Item27, Item39
8	Personal Adjustment	3	Item1, Item25, Item40
9	Perceived Instructor Efficacy	5	Item22, Item24, Item28, Item35, Item36
10	Socializing	3	Item13, Item17, Item37
	Total	49	

To ensure the quality of the instrument, a pilot study had been conducted on Form 2 students in

Sekolah Menengah Sains Tapah, Perak. The pilot study was conducted in order to improve the

instrument based on the analysis results, before it was administered to the respondents in the actual study. The pilot results showed that all the items were sufficient in quality and could be utilised for the actual study.

#### 4. Data analysis

The Kaiser-Meyer-Olkin (KMO) test (Table 5) with a value of 0.85 indicates that the sample is sufficient for the factor analysis test. With Bartlett's test showing a significant *chi-square* value ( $\chi^2=10161.02$ ,  $p<0.05$ ), this meant that the factor analysis test was appropriate and valid to be conducted. Instrument soundness was examined using principal components factor analysis with varimax rotation.

**Table 5:** KMO and Bartlett's test

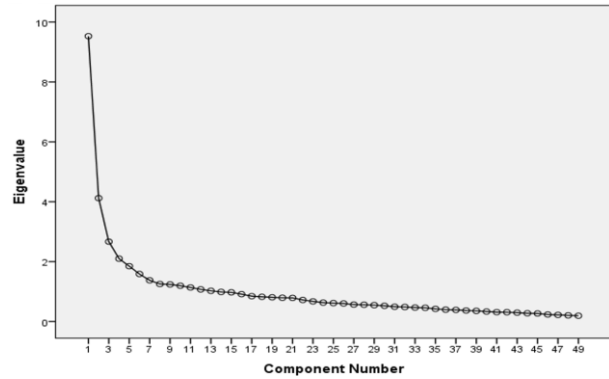
Kaiser-Meyer-Olkin Measure of Sampling adequacy.	0.85
Approx. Chi-Square	10161.02
Bartlett's Test of Sphericity	df 1176
Sig.	0.00

Before the data was analysed with an IRT-based software, two assumptions had to be fulfilled. Hishamuddin and Siti Eshah (2016) found that the unidimensionality and local independence assumptions should be tested before conducting an IRT-based analysis. As such, the *exploratory factor analysis* (EFA) was utilised to test the compatibility of unidimensional structures with the data and subsequently testing the local independence of items.

From Fig. 1, we can see that the first eigenvalue was much greater than the others. Therefore, it suggests that a unidimensional model is reasonable for this data (Hishamuddin and Siti Eshah, 2016). Hambleton et al. (1991) stated that, when unidimensionality assumption is met, then the local independence is also obtained. Since the unidimensionality assumption of the latent trait measured in this study is considered reasonable, therefore the assumption of local independence is also accepted.

Based on the EFA result, analysis output showed that the instrument constructs contributed 61.53% of the variance explained. Hair et al. (2010) stated that the accepted minimum value of total variance explained in the factor analysis is 60 percent. This indicated that the constructs in the study had sufficient construct validity.

The data analysis with -2 *LogLikelihood* (-2LL) statistics (Table 6) as proposed by de Ayala (2009) showed that the *Samejima's Graded Rating Model* (SGRM) was more suitable with the data presented as compared to the *Graded Rating Scale Model* (GRSM). The output was in line with the study by Demirtaş et al. (2016), which stated that the *Graded Rating Model* (GRM) showed a better model fit with polytomous data. Therefore, the data analysis was conducted based on the SGRM polytomous IRT model.



**Fig. 1:** Study data scree plot

**Table 6:** Comparison of -2LL statistics for the Selection of IRT Model

Model	Items	$\chi^2$	df	p	-2LL
GRSM	49	4245.015	3822	0.000	43125
SGRM	49	3909.234	3871	0.330	42775

## 5. Results

### 5.1. Instrument reliability

In research, the value of  $\alpha > 0.7$  is frequently referred as the 'cut-off value', 'minimum value', or 'good' for reliability index. However, Taber (2018) found that the value of  $\alpha \geq 0.45$  is categorized as 'acceptable' or 'sufficient' to prove the reliability or internal consistency of an instrument. Griethuijsen et al. (2015) in their study to measure students' interest towards science in selected countries found a few constructs with  $\alpha$  under the value of 0.7 or 0.6.

However, this study findings (Table 7) showed that the instrument reliability ( $\alpha=0.89$ ) was very good and exceeded the minimum value which was often used as the reference in some researches.

**Table 7:** Reliability of ASICS instrument

Construct	Number of items	Alpha ( $\alpha$ )
Full test	49	0.89

### 5.2. Instrument construct validity

In the context of polytomous IRT, the instrument validity could be assessed using item statistics. According to Guyer and Thompson (2013), *chi-square* statistics comprise an overall index showing how well the response data corresponds to the chosen IRT model. The *chi-square* statistics could be utilized for both dichotomous and polytomous items. For polytomous items, the *chi-square* value could be used to show items which do not fit or misfit. A *chi-square*  $p$  value which is less than 0.05 ( $p < 0.05$ ) would mean that the item does not fit the model. In other words, an item which shows a *chi-square*  $p$  value less than 0.05 ( $p < 0.05$ ) is an item which does not measure the construct properly. According to the rule of thumb, if most of the items (more than 70%) fit the model, then the construct validity is very good. In this study, it was found that 42 out of 49 or 85.7 percent (0.86) of items (Table 8) fit the model. As such, it could be stated that the ASICS instrument



used for this study had measured what it was supposed to measure very well.

**Table 8:** Chi-Square value for items

Q	$\chi^2$	p	Q	$\chi^2$	p	Q	$\chi^2$	p	Q	$\chi^2$	p
1	75.	0.	1	60.	0.	2	95.	0.	4	95.	0.
2	132	0.	1	96.	0.	2	126	0.	4	88.	0.
3	95.	0.	1	96.	0.	2	53.	1.	4	97.	0.
4	50.	0.	1	131	0.	3	63.	0.	4	93.	0.
5	63.	0.	1	35.	0.	3	33.	1.	4	67.	0.
6	57.	0.	1	43.	0.	3	104	0.	4	62.	0.
7	36.	0.	2	100	0.	3	66.	0.	4	74.	0.
8	44.	1.	2	100	0.	3	51.	1.	4	74.	0.
9	95.	0.	2	109	0.	3	54.	0.	4	118	0.
1	74.	0.	2	67.	0.	3	105	0.	4	64.	0.
1	60.	0.	2	128	0.	3	91.	0.	-	-	-
1	50.	1.	2	89.	0.	3	62.	0.	-	-	-
1	120	0.	2	81.	0.	3	66.	0.	-	-	-

### 5.3. Discrimination parameter

Discrimination parameter in the IRT context is also known as ' $a$ ' parameter. According to Baker (2001), the interpretation of the magnitude of discrimination of an item or construct is based on the output of  $a$  parameter. Guyer and Thompson (2013) stated that an item with higher discrimination parameter value was considered better than an item with lower discrimination parameter value. This study also looked into the value of  $a$  parameter as shown in a study by Mokshein (2018) which had applied the *Xcalibre* software for the item calibration process. She stated that, the value of discrimination parameter referred to and considered as more practical when it was similar or more than 0.3 ( $\geq 0.30$  or 0.3 to 1.5). If an item had not been discriminated ( $a < 0.3$ ), then the probability to acquire an accurate response would not increase much in line with the  $\theta$  increase of respondents (DeMars, 2010).

As shown in Table 9, most items of this study (45 out of 49) or 91.8 percent were discriminated well.

**Table 9:** Value of item discrimination parameters

Q	$a$	Q	$a$	Q	$a$	Q	$a$
1	0.25	14	0.71	27	0.35	40	0.38
2	0.43	15	0.56	28	0.13	41	0.70
3	0.31	16	0.73	29	0.66	42	0.56
4	0.71	17	0.48	30	0.89	43	0.69
5	0.76	18	0.57	31	0.90	44	0.90
6	0.73	19	0.63	32	0.36	45	0.97
7	0.49	20	0.23	33	0.74	46	0.66
8	0.66	21	0.63	34	0.91	47	0.61
9	0.78	22	0.43	35	0.74	48	0.40
10	0.68	23	0.74	36	0.44	49	0.74
11	0.65	24	0.43	37	0.46		
12	0.59	25	0.24	38	0.46		
13	0.40	26	0.37	39	0.52		

### 5.4. Theta score on constructs

Based on the traditional procedures for calculating ASICS scores, Prevatt et al. (2011) found that important factors in predicting grade point

average (GPA) or academic success were *Personal Adjustment*, *General Academic Skills*, *Internal Motivation / Confidence*, *Socialization*, and *Concentration*.

However, the score in this study was based on the theta parameter or  $\theta$  estimate, which can be obtained from the data analysis using *Xcalibre* software. The  $\theta$  score described the level or abilities of the respondents towards the measured constructs. Although the respondents could be categorized as high-performing students, by using IRT, the researchers could further categorize them as high, medium, and low ability students.

According to Thompson (2009), the  $\theta$  score at 0.0 could refer to the value of medium ability at the  $\theta$  scale. Respondents who had a  $\theta$  score less than 0.0 could be considered as low ability students while those who had a  $\theta$  score more than 0.0 could be considered as high ability students (Guyer and Thompson, 2013). Students who categorized as 'low' in the context of this study were actually high achieving students. But, they have less ability depending on the studied constructs as compared to the other respondents in this study.

Based on the theta score for every construct as shown in Fig. 2, it was found that the '*Perceived Instructor Efficacy*' and '*Personal Adjustment*' were two factors chosen by the high-ability students. This meant that the students categorized as having high abilities had chosen '*Perceived Instructor Efficacy*' and '*Personal Adjustment*' as important factors, which helped to propel their academic success.

## 6. Discussion

### 6.1. Instrument quality

Originally, Prevatt et al. (2011) in their study had reported the ASICS internal consistency according to each constructs. The results of this study are in line with the study conducted by Prevatt et al. (2011) where *General Academic Skills* construct show the highest internal consistency, while the *External / Current Motivation* construct shows the lowest internal consistency.

However, Prevatt et al. (2011) did not report the overall internal consistency value of the instrument. This study found that, overall internal consistency in which ASICS instruments are administered is very high as stated with  $\alpha = 0.89$ . The value is interpreted as very high based on studies by Griethuijsen et al. (2015) and Taber (2018).

The construct validity for the instrument was also found as very high which 86.0 percent of the items had measured what it was supposed to measure as well as most items (91.8%) showed good discrimination.

### 6.2. The highest factors scored based on theta

Prevatt et al. (2011) in their research which based on their ASICS scoring procedures found that

*Personal Adjustment, General Academic Skills, Internal Motivation / Confidence, Socializing, and Concentration* were the most important factors contributed in academic success, it is not necessarily the same factors researchers will get in other research setting. Using the same traditional scoring

procedures as introduced by [Prevatt et al. \(2011\)](#), this study found that *External Motivation / Future* and *External Motivation / Current* found to be the highest ASICS scored factors and considered as important role in academic success.

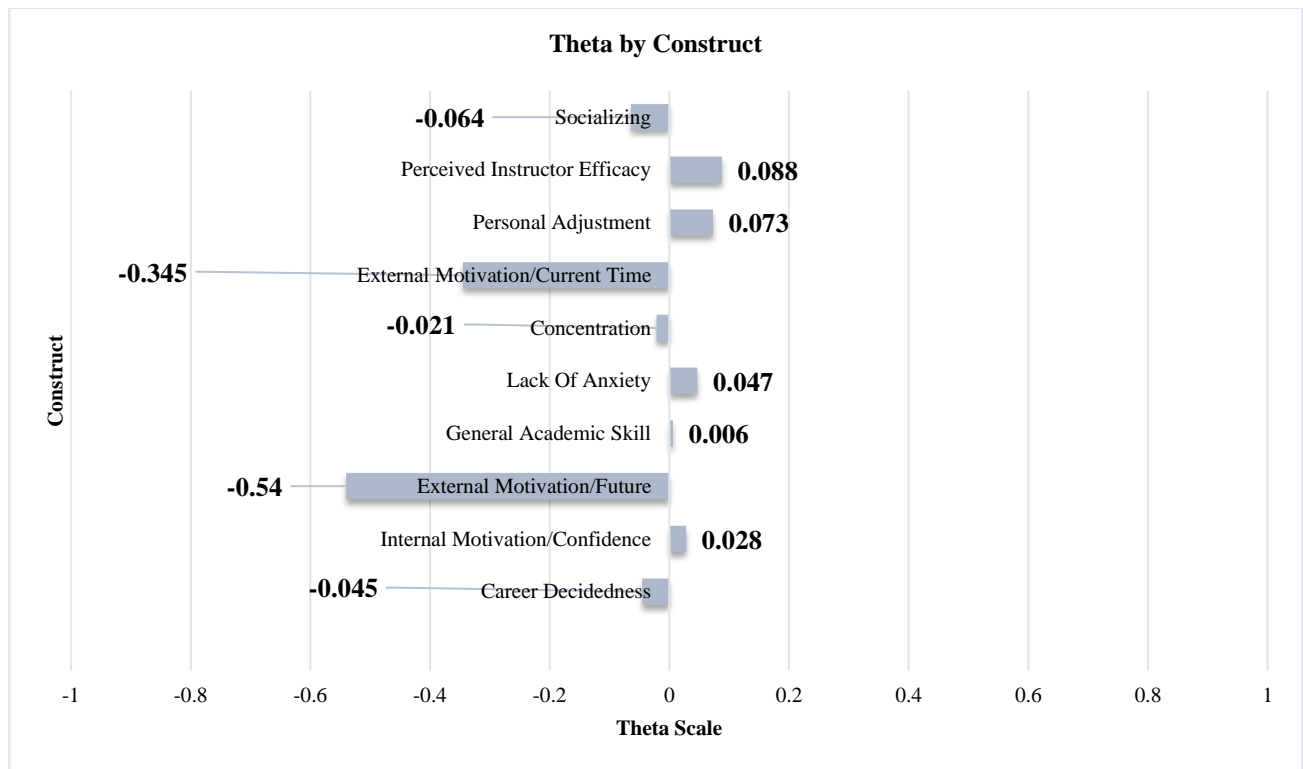


Fig. 2: Respondents' theta according to construct

It shows, the contributed factors in academic success were different between the two studies. Furthermore, the factors are all showing a global picture of overall students' performance. That means these factors are pictured by the whole group of respondents without further classifying them into high, moderate, and low ability.

Further analyses with the application of polytomous IRT and the concept of theta scales, this study also found that the highest ability group of students are more prefer to *Perceived Instructor Efficacy* and *Personal Adjustment* factors, whereas the lowest ability group of students need more *External Motivation/ Future* and *External Motivation/ Current* for their academic success. This study shows where students which at the highest ability among the high performance group of students are more likely to choose *Perceived Instructor Efficacy* and *Personal Adjustment* as important factors in their academic success.

## 7. Conclusion

The ASICS was found to be a reasonably sound instrument for measuring academic success among students in FRS. *Perceived Instructor Efficacy* subscale is found as the highest scored construct among the highest performance group of students. This study also found that the highest performance

students are personally adjusted by their high performance school surroundings for their academic success.

The findings show that the role of effective teachers is very important in the student's performance and development. It means, with the development in teaching and learning scenarios, the aspect of giving too much empowerment to students in organizing their own studying or managing their own learning should be managed properly. Perception towards teachers' effectiveness are always important for some students. Aspect which should also be focused on is related to the role of the teachers, where they should not remain passive in teaching and become dependent on educational technology although we are in the *Industrial Revolution 4.0* (IR4.0) era. It means, to develop an excellent future generation, we still require an efficient education sector, especially one which focuses on efficient teachers for the 4.0 Industrial Age.

## 8. Recommendations

The recommendations reached through this research are as followed:

- Although IRT polytomous does not substitute Classical Test Theory (CTT) or classical methods

that were applied worldwide especially in the psychometric field, IRT polytomous is a recommended and worthwhile analyses tool that can and should be used to further analyse the quality of psychometric instrument.

- In order to develop the success and high performance future generations, developing educational competencies (teachers) for education in the IR4.0 era is a first thing to do as well as creating and maintaining a conducive environment to support teaching and learning.

## Acknowledgment

The authors wish to express our sincere appreciation to Research Management and Innovation Centre (RMIC), Universiti Pendidikan Sultan Idris for the research grant funding (Research Code: 2017-0276-107-01). Also greatly appreciations to the teachers and students studying at FRS schools in Seremban district namely Kolej Tunku Kurshiah, Sekolah Menengah Sains Tuanku Munawir, and Sekolah Dato' Abdul Razak who assisted in data collection.

## Compliance with ethical standards

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Baker FB (2001). The basics of item response theory. Education Resources Information Center: ERIC, College Park, USA.
- Cohen RJ, Swerdlik ME, and Smith DK (1992). Psychological testing and assessment: An introduction to tests and measurement. 2<sup>nd</sup> Edition, Mayfield Publishing Company, Mountain View, USA.
- de Ayala RJ (2009). The theory and practice of item response theory. Guilford Press, New York, USA. **PMid:19590082**
- DeMars C (2010). Item response theory. Oxford University Press, Oxford, UK.  
<https://doi.org/10.1093/acprof:oso/9780195377033.001.0001>
- Demirtaş RN, Yalçın S, and Cansu AYAN (2016). The development of IRT based attitude scale towards educational measurement course. Journal of Measurement and Evaluation in Education and Psychology, 7(1): 133-144.  
<https://doi.org/10.21031/epod.43804>
- Ghani AFM, Siraj S, Kassim R, Kenayathulla HB, Marzuki S, and Elham F (2013). Amalan sekolah cemerlang di sekolah berasrama penuh dan sekolah menengah kebangsaan agama: Satu perbandingan. Online Journal of Islamic Education, 1(2): 31-50.

- Griethuijsen RA, van Eijck MW, Haste H, den Brok PJ, Skinner NC, Mansour N, and BouJaoude S (2015). Global patterns in students' views of science and interest in science. Research in Science Education, 45(4): 581-603.  
<https://doi.org/10.1007/s11165-014-9438-6>
- Guyer R and Thompson NA (2013). User's manual for Xcalibre™ item response theory calibration software, version 4.2. Assessment Systems Corporation, Woodbury, Burbank, USA: 44-76.
- Hair JF, Black WC, Babin BJ, and Anderson RE (2010). Multivariate data analysis. Pearson Education, London, UK.
- Hambleton RK, Swaminathan H, and Rogers HJ (1991). Fundamentals of item response theory. Sage Publications, Newbury Park, USA.
- Hishamuddin A and Siti Eshah M (2016). Is 3PL IRT an appropriate model for dichotomous item analysis of AandP final exam?. Malaysian Science and Mathematics Education Journal, 6(1):13-23.
- KPM (2017a). Bahagian Pengurusan Sekolah Berasrama Penuh dan Sekolah Kecemerlangan. Kementerian Pendidikan Malaysia, Kuala Lumpur, Malaysia. Available online at: <https://www.moe.gov.my/>
- KPM (2017b). Perangkaan Pendidikan Malaysia. Bahagian Perancangan dan Penyelidikan Dasar Pendidikan, Putrajaya, Malaysia: 26-56. Available online at: <https://www.moe.gov.my/>
- KPM (2018). Portal Rasmi Jabatan Pendidikan Negeri Sembilan. Kementerian Pendidikan Malaysia, Kuala Lumpur, Malaysia. Available online at: <https://www.jpnn.moe.gov.my>
- Krejcie RV and Morgan DW (1970). Determining sample size for research activities. Educational and Psychological Measurement, 30(3): 607-610.  
<https://doi.org/10.1177/001316447003000308>
- MOE (2018). Bahagian Perancangan dan Penyelidikan Dasar pendidikan. Ministry of Education Malaysia, Kuala Lumpur, Malaysia. Available online at: <https://www.moe.gov.my/>
- Mokshein SE (2018). Aplikasi teori respons item dalam penyelidikan. Universiti Pendidikan Sultan Idris Publisher, Malaysia.
- Ostini R and Nering ML (2006). Polytomous item response theory models. Sage, Thousand Oaks, USA.  
<https://doi.org/10.4135/9781412985413>
- Prevatt F, Li H, Welles T, Festa-Dreher D, Yelland S, and Lee J (2011). The academic success inventory for college students: Scale development and practical implications for use with students. Journal of College Admission, 211: 26-31.
- Rivers DC, Meade AW, and Lou Fuller W (2009). Examining question and context effects in organization survey data using item response theory. Organizational Research Methods, 12(3): 529-553.  
<https://doi.org/10.1177/1094428108315864>
- Taber KS (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education, 48(6): 1273-1296.  
<https://doi.org/10.1007/s11165-016-9602-2>
- Thompson NA (2009). Ability estimation with IRT. White Paper, Assessment System Corporation, Minnesota, USA.