International Journal of Psychosocial Rehabilitation, Vol.24, Issue 09, 2020
ISSN: 1475-7192

Validation of The Academic Success Instrument Using Polytomous Item Response Theory

[1] Hishamuddin Ahmad, [2] Siti Eshah Mokshein, [3] Mohd Razimi Husin, [4] Ismail Yusuf Panessai, [5] Mohd Nazir Md Zabit[a], [6] Nor Hasnida Che Md Ghazali

Abstract

Several instruments that measuring an academic success have been developed and published. However, a psychometrically sound instrument to measure academic success that suitable for the Malaysian context needs to be validated. Therefore, this study aimed to validate on Academic Success Inventory for College Students (ASICS) instrument which contains 49 items in Malaysian context. The sample comprised 305 students from three schools. Data was analyzed based on the Polytomous Item Response Theory (IRT) 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 most 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 (α=0.89) and the construct validity was also fulfilled with the value of 0.86. Therefore, this instrument adds to the limited collection of locally validated instruments in the field of educational assessment and evaluation.

Keywords: *validation, academic success instrument, polytomous IRT.*

Introduction

There are several instruments that measuring an academic success have been developed and published. John, Greg, Linda, Elizabeth, and Karen (2014) stated that academic success skills are prerequisites to content learning. Travis, Charles, and Susan (2015) defined what constitutes student success. They have found that there is no complete presentation of empiric instruments available to educational researchers seeking to measure various aspects of academic success. Based on their findings, academic success was defined as inclusive of academic achievement, attainment of learning objectives, acquisition of desired skills and competencies, satisfaction, persistence, and post college performance. Sadeghi-Gandomani and Adib Hajbaghery (2018) in their study used an instrument that contained 10 subscales that assess important aspects of academic success.

However, this study is focused on Academic Success Inventory for College Students (ASICS) instrument. ASICS' development was based on a large public university in the southeastern United States university in the southeastern United States (Prevatt, Li, Welles, Festa-Dreher, Yelland, & Lee,, 2011). A psychometrically sound instrument to measure academic success that suitable for the Malaysian context needs to be validated. Therefore, this study aimed to validate on ASICS instrument for secondary students in Malaysia.

Polytomous Item Response Theory (IRT) is applied in validating the instrument. The polytomous IRT model is generalized from the dichotomous IRT model when more than two categories exist. In other words, the polytomous model is for items that are not scored binarily or wrongly / true (1/0). When some items in the test are scored with more than two response categories such as Likert scale, a polytomous IRT model is required instead of dichotomous IRT model. Ostini and Nering (2006) stated that the simplest and most obvious reason for the development of the IRT polytomous model is

[1] *Faculty of Human Development, University Pendidikan Sultan Idris, Malaysia.*
Email: hishamuddin.a@fpm.upsi.edu.my
[2,3,4,5,6] *Faculty of Human Development, University Pendidikan Sultan Idris, Malaysia*

the fact that polytomous items are widely used and usually applied in the field of psychological measurement. They also highligted that all responses in the category or Likert scale can be analyzed using this model of polytomous.

There are several polytomous models that exist. Each model illustrates the extent to which IRT basic philosophy for non-binary data. Each polytomous model determines the behavior of candidates as their latent trait function (often known as ability). Ordered category polytomous items are items in which the response categories have a clear rank associated with the nature of the study. Likert scale items and partial credit cognitive abilities are examples of polytomous items in the form of ordered categories. In literacy, such items are also known as graded responses. This study aims to validate an academic success instrument. Particularly, the study objectives are to assess reliability and construct validity of the instrument using Polytomous Item Response Theory.

Methodology

This is a quantitative study with a survey method. It applies IRT polytomous model in responding to research questions. Respondents (305) from a population of 389 secondary school students were sampled. They are from three secondary schools in Seremban (Table 1). In the context of IRT, a sample which closely resembles the actual population in terms of numbers is preferred to describe the findings of the study (Embretson & Reise, 2000). DeMars (2010) stated 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 in this study (**Table 1**) is considered adequate to make a generalization of the population in this study.

Table 1. Population and study sample from three schools

No.	Schools	Population	Sample
1.	School A	143	105
2.	School B	119	109
3.	School C	127	91
	Total	389	305

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*.

To use the instrument in conducting this study, permission was granted to use the measure on May 20, 2014. Although Cohen and Swerdlik (2002) suggest 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 Malaysian 'climate'.

According to the expert who had evaluated the instrument, one item related to 'drink', which referred to an alcoholic drink had been excluded. 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 2**.

Table 2. Constructs and items

Code	Construct	No. of items	Items
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	

Data Analysis

The *Kaiser-Meyer-Olkin* (KMO) test 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.

Before the data was analysed with an IRT-based software known as *Xcalibre*, 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 scree plot output, the first eigenvalue was found much greater than the other eigenvalues. Therefore, it suggests that a unidimensional model is reasonable for this data (Hishamuddin & 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.

The data analysis with *-2 LogLikelihood* (-2LL) statistics 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 Demirtaşlı, Yalçın, and Ayan (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.

Results

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 (2017) found that the value of $\alpha\geq0.45$ 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 α under the value of 0.7 or 0.6. However, this study found 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.

Instrument Construct Validity

Validity and reliability are the important attributes for the quality of an assessment. Kelley (1927) stated that the problem of validity concerns with whether a test really measures what it purports to measure while reliability is the question on

how accurately a test measures the thing which it does measure. In the context of polytomous IRT, the instrument validity could be assessed using *chi-square* 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 instrument construct validity is very good. In this study, it was found that 42 out of 49 or 85.7 percent (0.86) of items fit the model. As such, it could be stated that the ASICS instrument used in this study had measured what it was supposed to measure very well.

Conclusion

This study found that, internal consistency for ASICS instruments is very high as stated with α=0.89. 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. This means, the instrument can be used appropriately to measure the academic success among students in Malaysia so that appropriate follow-up actions can be implemented towards the betterment of students' education quality.

References

Cohen, R. J., & Swerdlik, M. E. (2002). *Psychological Testing and Assessment: An Introduction to Test and Measurement* (5th ed.). McGraw-Hill Companies, Inc.

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* New York: The Guilford Press.

Demirtaşlı, N., Yalçın, S., & Ayan, C. (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.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Griethuijsen, R. v., Eijck, M. v., Haste, H., Brok, P. d., Skinner, N. C., Mansour, N., . . . BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in Science Education, 45*(4), 581–603. doi:10.1007/s11165-014-9438-6

Guyer, R., & Thompson, N. A. (2013). *User's Manual for Xcalibre™ Item Response Theory Calibration Software, Version 4.2.* Woodbury MN: Assessment Systems Corporation.

Hambleton, R. K., Rogers, H. J., & Swaminathan, H. (1991). *Fundamentals of Item Response Theory.* USA: Sage Publication.

Hishamuddin Ahmad; Siti Eshah Mokshein. (2016). Is 3PL IRT an appropriate model for dichotomous item analysis of A&P final exam? *Malaysian Science & Mathematics Education Journal, 6*(1), 13-23.

John, C., Greg, B., Linda, W., Elizabeth, V., & Karen, H. (2014). Development of an instrument to measure student use of academic success skills: An exploratory factor analysis. *Measurement and Evaluation in Counseling and Development, 47*(3), 171-180. doi:10.1177/0748175613505622

Kelley, T. L. (1927). *Interpretation of Educational Measurements.* New York: World Book Company.

Ostini, R., & Nering, M. L. (2006). *Polytomous Item Response Theory Models.* Thousand Oaks, California: Sage Publications, Inc.

Prevatt, F., Huijun, L., Welles, T., Festa-Dreher, D., Yelland, S., & Lee, J. (2011). The academic success inventory for college students: Scale development and practical implications for use with students. *Journal of College Admission*, 26-31.

Sadeghi-Gandomani, H., & Adib-Hajbaghery, M. (2018). Psychometric properties of Persian version of academic success inventory for college students. *Nursing and Midwifery Studies, 7*(4), 174-179.

Taber, S. K. (2017). The use of Cronbach's alpha when developing & reporting research instruments in science education. *Research in Science Education*, 1-24. doi:10.1007/s11165-016-9602-2

Travis, T. Y., Charles, G., & Susan, R. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation, 20*(5), 1-20.