The Effects of Sample Size and Options Number on the Validity Item of Students' Environmental Personality Score

Rahmirini Datau[1,*] I M. Putrawan[2,] Wardani Rahayu[3]

[1, 3] *Educational Research and Evaluation, State University of Jakarta*
[2] *Environmental Education and Management, State University of Jakarta*
[*]*Corresponding author. Email*: rahmirinidatau0@gmail.com

ABSTRACT

Personality possesses close ties with character education or building. A student that studied character building will be able to apply pro-environmental behaviors. An instrument is needed to determine a students' environmental personality, and the most important trait in a test/instrument is validity. Item validity is used to measurement to improve it, where accurate measurement is obtained not only through item accuracy but also through the effects of the options number and sample size. This study aims to uncover the difference between the item validity of students' environmental personalities with the options number and sample size. The data are analyzed by using a two-way ANAVA with Tukey post-hoc test. The results of this study show that (1) Students environmental personality's item validity is affected by the number of options, (2) the item validity of students' environmental personality is not affected by sample size, (3) There is an interacting influence between the number of options and sample size on the item validity of the students' environmental personality score. Based on the Tukey test we discovered that the group working in 3 options with a sample size of 175 students, differed compared to other groups significantly. In conclusion, the highest item validity for students' environmental personality test is achievable on the group that uses three options with a sample size of 175 students.

Keywords: *Item validity, Environmental personality, Number of options, Sample size.*

1. INTRODUCTION

Awareness to reduce the impact of environmental destruction and climate change is still incredibly low among the general population. The existence of pro-environmental behavior within an individual may be proven through actions such as conserving water, reducing electricity usage, using public transportation, and other activities that protect the environment [1].

Protecting the environment is one of the main challenges of modern society, and therefore it is important to understand the pro-environmental behavior within society at large [2], especially students, through an educational institution. Education possesses close ties with the concept of character education, this is because characteristics and traits in an individual make up their personality, which is a set of values that set in stone the way an individual think, act, and behave in such a way that it becomes the identity of an individual that is shaped by their habits. [3]. One of the most important character traits that should be taught to students at an early age is pro-environmental behavior. This trait values actions that would prevent damaging the surrounding environment.

A student who has studied character education will be able to apply pro-environmental behavior in their day-to-day life. Protecting the environment is done by applying environmental education at a young age, educating them through teachers, and raising environmental awareness. To measure students' environmental personality an instrument is needed, and one of the most important traits in an instrument is validity.

Validity is an incredibly important trait for a test because to measure psychological attributes accurately, high validity is needed [4]. To make items that contribute as little measurement error as possible, measurements and various research methods cover response scales that restrict and regulate the choices that are available for the respondents.

Studies in psychology and other fields have always involved issues regarding various types of measurement scales and the number of available options

Given the importance of score variance in classical test theory, a response scale that increases the variance of valid scores should increase the precision of the measurement, thereby maximizing the coefficients of validity and reliability [5]. Regardless of how response scales perform on psychological measurements, environmental personality scales have been expanded and elaborated in various ways ever since their inception. Even though sources regarding the matter are few and inconsistent, there have been several studies regarding the nature of option sizes, their response, and respondents' psychological behavior. The data that will be measured commonly decides the choices made by researchers regarding scale responses [6].

As of now, the optimum number of options remain a highly argued topic, this is caused due to a lack of proper conclusions from theoretical and empirical findings. One such argument regards the use of three, four, or five options to reduce the impact of respondents' random guesses. Thus far the use of five options remains to be the most common number of options used, but more studies suggest the use of three options due to it possessing similar psychometric effects, as well as being more efficient in terms of research management and development.[7].

There are several schools of thought that may affect the views of an individual regarding the use of the number of options in an environmental personality scale. First, one must take into account the importance of score variants in a classical test theory. Second, a response scale that produces an increase of score variance must increase measurement precision and thereby maximize the reliability and validity coefficient.

Within measurement there are two types of error, they are sampling error and systematic error. Sampling error is the difference between the true score of the population and the obtained data. This is caused due to the sample size which is merely a fraction of the possibilities that are obtainable from the total sum of the population. Sampling error remains a possibility regardless of the differences in measurement, situation, and condition during measurement, as well as the data, being measured [8].

There are various discussions regarding sample size, but in this study, we took the middle ground by comparing sample sizes of less than 200, and more than 200 people. According to [7] at least a sample of 200 respondents is needed for stability. Many studies show the impact of sample sizes on psychometric estimation [9]. But so far recent studies have only discussed measuring item validity of environmental personality, the number of samples, and sample sizes individually or in a different context. But there has been no study so far that explored the interaction of each variable by comparing less than 200 and more than 200 samples, as well as the effects of having three options compared to five, using two-way ANAVA with a 2x2 design, therefore this study possess a novelty. Furthermore, this study aims to find the cause of the rise and decline of the pro-environmental personality measurement's item validity that is caused by the number of options and sample size.

2. MATERIALS AND METHODS

This research was conducted at SMAN 16, Jakarta. The time of the research was carried out in the Odd Semester of 2021/2022. The research sample was students of SMAN 16 Jakarta. This study aims to discover the effects of the number of options and sample sizes on item validity; to achieve that, this study uses two-way ANAVA. Table 1 showed the design analysis that is used in this research.

Table 1. 2 x 2 Design analysis

Sample Size (B)	Number of Options (A)	
	3 options (A1)	5 options (A2)
Less than 200 students	A_1B_1	A_2B_1
More than 200 students	A_1B_2	A_2B_2

Description:

A_1B_1 = Average item validity of students' personality score with three options with a sample size of less than 200 students (175).

A_2B_1 = Average item validity of students' personality score with five options with a sample size of less than 200 students (175).

A_1B_2 = Average item validity of students' personality score with three options with a sample size of more than 200 students (215).

A_2B_2 = Average item validity of students' personality score with five options with a sample size of more than 200 students (215).

3. RESULTS AND DISCUSSION

The data is analyzed with the Kolmogorov Smirnov normality test and Levene's homogeneity test to discover the normal distribution of the data obtained. Then an analysis of the two-way ANAVA test with a post hoc Tukey test was used to determine the effects of the number of options and sample size on item validity. Table 2 showed the results of the normality test.

The results of the Kolmogorov Smirnov normality test shows that the significance score of the four groups is ($p > 0.05$) therefore the data possess normal distribution. The obtained data is further analyzed with Levene's homogeneity test.

The results of Levene's test show a significance score = $0.051 > 0.5$ therefore this variant of the population is homogenous. The data were analyzed with a two-way ANAVA test. Table 4 showed the result of the two-way ANAVA test. By reviewing the item validity of the students' environmental personality score based on the number of options on the level 5% we can compare between F count = $10.93 >$ F table = 3.34 with a significance of $0.001 < 0.05$ in which H_0 is rejected, therefore there is a significant difference on the item validity of the students' environmental personality score affected by the number of options. This result is in line with a study that having 3 to 5 options are the most common response used and both options can increase psychometric qualities, including reliability and validity [10].

By reviewing the item validity of the students' environmental personality score based on sample size on the level 5% we can compare between F count = $1.423 <$ F table = 3.34 with a significance $0.235 > 0.05$ in which H_0 is accepted, therefore there is no significant difference on the item validity of the students' environmental personality score affected by the sample size. Sampling error is the difference between the true score of the population and the data

obtained within the research. This occurs because the data and measurements obtained within the populations are merely one of the many possibilities of the data that can be obtained from the said population. Table 5 showed the result of the Turkey test.

Sampling error remains a possibility regardless of the differences in measurement, situation, and conditions during measurement, as well as the data being measured [11].

By reviewing the interaction between the number of options and sample size on item validity of the students' environmental personality by the level of 5% we can obtain F count = $10.936 >$ F table = $3. 34$ with a significance of $0.001 < 0.05$ which means H_0 is rejected, therefore there is interacting influence between the number of options and sample size on the item validity of students' environmental personality.

The data analyzed through the ANAVA test is then analyzed with the Honestly Significance Difference (Tukey) post hoc test to analyze the difference between each group.

The results of the Tukey test shows that there is a meaningful difference between the groups. The group with 3 options and a sample of 175 students are significantly different compared to other groups, while other groups such as 3 options and sample of 215 students, 5 options and a sample of 175 students, and 5 options and a sample of 215 students show no significant difference to one another. Therefore, we concluded that for item validity of students' environmental personality, working with 3 options and a sample of less than 200 is better compared to working with other combinations of options and sample size. This result is also supported by the finding of [12]. That supports using 3 options due to it offering some advantages, such as higher content validity and reliability. This is also in line with a study done by [13]. which states for an instrument that uses 5 options per item needs at least 250

Table 2. Results of the normality test

Tests of Normality		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
interaction		Statistic	df	Sig.	Statistic	df	Sig.
Item Validity	option 3 samples 175	.099	30	.200*	.965	30	.404
	option 5 sample 175	.133	30	.189	.933	30	.058
	option 3 sample 215	.115	30	.200*	.947	30	.140
	option 5 sample 215	.103	30	.200*	.951	30	.185

Table 3. Results of Levene's test

Levene's Test of Equality of Error Variances[a]			
Dependent Variable: item validity			
F	df1	df2	Sig.
2.668	3	116	.051

Tabel 4. Results of two-way ANAVA test

Tests of Between-Subjects Effects						
Dependent Variable: item validity						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.406[a]	3	.135	7.485	.000	.162
Intercept	8.641	1	8.641	478.535	.000	.805
option	.191	1	.191	10.553	.002	.083
sample	.025	1	.025	1.368	.244	.012
option * sample	.190	1	.190	10.535	.002	.083
Error	2.095	116	.018			
Total	11.142	120				
Corrected Total	2.500	119				

Tabel 5. Results of Tukey test

Pairwise Comparisons						
Dependent Variable: item validity						
(I) interaction	(J) interaction	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
option 3 samples 175	option 5 samples 175	-.159*	.035	.000	-.228	-.091
	option 3 samples 215	-.108*	.035	.002	-.177	-.040
	option 5 sample 215	-.108*	.035	.002	-.177	-.040

item requires at least 100 respondents.

4. CONCLUSION

The number of options can influence the item validity of students' environmental personality score, this is proven with F count =10.93 > F table = 3.34 with a significance of 0.001 < 0.05 which means H_0 is rejected. Meanwhile, sample sizes do not appear to influence the item validity of students' environmental personality score, this is proven with F count = 10.936 > F table = 3. 34 with a significance of 0.001 < 0.05 which means H_0 is rejected. Tukey's post hoc test shows that the group of samples that uses 3 options with a sample of 175 students is significantly different compared to other groups, while other groups are not significantly different from one another. Therefore, the item validity of students' personality tests in the group working with 3 options with 175 students is much better compared to other options and sample size combinations.

Further research on pro-environmental personality should expand on this research by using 3 options on a greater variety of sample sizes. This study can hopefully serve to help other researchers in the subject of environmental personality, especially their connection with the number of options used and sample size in measurement. A more accurate instrument for measuring students' environmental awareness may also help teachers in measuring students' demeanor in line with the current character education program, as environmental personality and awareness could also serve as an indicator in character education.

AUTHORS' CONTRIBUTIONS

All authors conceived and designed this study. All authors contributed to the process of revising the manuscript, and at the end, all authors have approved the final version of this manuscript.

REFERENCES

[1]   M. A. Vicente-Molina, A. Fernández-Sáinz, J. Izagirre-Olaizola, Environmental knowledge and other variables affecting pro-environmental behaviour: Comparison of university students from emerging and advanced countries, Journal of Cleaner Production, 61 (2013) pp. 130–138.

[2]   I. M. Putrawan, Predicting environmental leadership based on personality and its construct validity, Internasional Jurnal Enginering Technology, 7(4) (2018) pp. 92–95.

[3]   D. C. Funder, Accurate Personality Judgment, Current Directions in Psychological Science 21(3) (2012) 177–182, doi: 10.1177/0963721412445309.

[4]   S. Azwar, Reliabilitas Dan Validitas Aitem, Buletin Psikologi, 3(1) (2016) pp. 19–26 doi: 10.22146/bpsi.13381.

[5]   L. J. Simms, K. Zelazny, T. F. Williams, L. Bernstein, Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data, Psychological assessment, 31(4) (2019) 557–566 doi: 10.1037/pas0000648.

[6]   J. A. Colquitt, J. A. Lepine, M. J. Wesson, Improving performance and outcomes, 4(4) (2005).

[7]   C. M. Suchindran, Sample Size Sample Size, Sampl. Choos. Cases Quality Research A Realistic Approach, 3(10) (2014) p. 2-6.

[8]   J. C. Hobart, S. J. Cano, T. T. Warner, A. J. Thompson, What sample sizes for reliability and validity studies in neurology?, Journal of neurology, 259(12) (2012)pp. 2681–2694, doi: 10.1007/s00415-012-6570-y.

[9]   P. Lietz, Research into questionnaire design: A summary of the literature, *I*nternational journal of market research, 52(2) (2010) p. 249, [Online]. Available: http://simidchiev.net/fid/Dissert/Survey/Researchintoquestionairedesign.pdf.

[10] J. Dawes, Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales, International journal of market research, 50(1) (2008) 61–77, doi: 10.1177/147078530805000106.

[11] T. Rijanto, Pengaruh Metode dan Ukuran Sampel Terhadap Variansi Skor Hasil Penyetaraan, Penelitian dan Evaluasi Pendidikan., 3 (2012) 365–383.

[12] M. S. Trevisan, G. Sax, W. B. Michael, Estimating the optimum number of options per item using an incremental option paradigm, Educational and Psychological

Measurement, 54(1) (1994) 86–91, doi: 10.1177/0013164494054001008.

[13] E. Hariadi, Penskalaan Butir Format Respons Pilihan dan Respons Bebas Berdasarkan Model Rasch dan Partial Credit, Jurnal Penelitan dan Evaluasi Pendidikan., 9(1) (2007), doi: 10.21831/pep.v9i1.1992.