Research Article doi: 10.12973/eu-jer.10.2.729

European Journal of Educational Research
Volume 10, Issue 2, 729 - 742.
ISSN: 2165-8714
https://www.eu-jer.com/

Assessment of the Validity and Reliability of Mental Health Instruments of High School Student in Indonesia

Heny Narendrany Hidayati*
Universitas Islam Negeri Syarif Hidayatullah, INDONESIA

Bahrul Hayat
Universitas Islam Negeri Syarif Hidayatullah, INDONESIA

Wardani Rahayu
Universitas Negeri Jakarta, INDONESIA

Received: April 8, 2020 ▪ Revised: December 27, 2020 ▪ Accepted: March 10, 2021

Abstract: This study aims to develop a standard instrument for measuring mental health among urban adolescents in Indonesia. The objective is to produce valid and reliable school adolescent mental health instruments to be used by agencies or schools to identify students' mental health. The survey was conducted in Jakarta and South Tangerang with a total of 1007 respondents divided into two experiments where the first trial was conducted on 597 students and the second trial was conducted on 410 students. Measurements were made using a Likert scale questionnaire. Instrument testing begins with a theoretical validity test by 4 experts and 20 panelists who test the instrument material in terms of construction, content and language. Experts analyze and correct the instrument qualitatively. The instrument was then reviewed and analyzed quantitatively by panelists using the Aiken index. At this stage, 44 items, 9 indicators and 3 variable dimensions were obtained. The next test is done by testing the validity empirically, by analyzing the measurement model using Confirmatory Factor Analysis (CFA) with the LISREL 8.80 Full Version program. By using the criteria for the SLF value ≥0.30 and t-value ≥1.96, and calculating the reliability with the construct reliability (CR) at the level> 0.70, the results of the second trial showed that 35 items were valid. The observations of the model fitness through Goodness-of-Fit test showed that there is a fitness between the theoretical model and the empirical model for the mental health instruments in this study.

Keywords: *Validity, reliability, mental health, item measurement, adolescence, Indonesia.*

To cite this article: Hidayati, H. N., Bahrul Hayat., & Rahayu, W. (2021). Assessment of the validity and reliability of mental health instruments of high school student in Indonesia. *European Journal of Educational Research, 10*(2), 729-742. https://doi.org/10.12973/eu-jer.10.2.729

Introduction

Mental health involves an individual's ability to feel comfortable with stress in their environment. The ability to respond to stress can be seen through thoughts, feelings, and behaviors that are appropriate to age, norms and local culture. They are also able to actualize their potential. Maslow (1981) claims that self-actualization is a psychologically healthy state, as the fulfillment of the highest needs (Townsend & Morgan, 2017). Because of its important role in measuring mental health, several measurements are widely used across a wide range of ages. However, Hammarström et al. (2016) highlighted the lack of validity measurement results for mental health problems among adolescents. For this reason, this study conducts research on the development of mental health instruments in the form of scales. This instrument is expected to be used as a first step to identify the mental health of students in religious schools or public educational institutions. The development of this instrument is focused on developing mental health instruments for high school students who are in the adolescent phase (Clauss-Ehlers et al., 2013; Omoniyi, 2016).

Good instruments generally are valid and reliable (Cohen et al., 1996; Linn, 2008). The instrument is said to be valid if it has accuracy in measuring the results, while the reliable instrument indicates the measurement results with a stable or consistent score (Reynolds et al., 2009). Fulfillment of the valid and reliable requirements of an instrument shows that the instrument has been standardized (Eryılmaz & Sapsağlam, 2018). This means that the instrument is developed through a rational validation process by experts and panels which is then refined according to the advice of experts and panelists. After the empirical validation process or trials, the instruments are then calculated, analyzed and assembled into a set of valid and reliable instruments (Kubiszyn & Borich, 1987; de Almeida Vieira & Fernandes, 2016).

* **Corresponding author:**
Heny Narendrany Hidayati. Universitas Islam Negeri Syarif Hidayatullah, Faculty of Education and Teacher Training, Indonesia.
✉ heny.fitk@yahoo.com

This study refers to the stage of developing standard instruments, in regards to the constructs of the measured variable of mental health. This study was conducted in Islamic high school, or *Madrasah Aliyah* (MA), students that are in their adolescence, i.e. in the transition period. Based on the existing problems, this study aims to produce a valid and reliable school adolescent mental health instrument, so that it can be used by the institution or school to identify the initial mental health of students and get a general picture of the mental health of the students (Liu et al., 2013). Students who are mentally healthy are characterized by having positive feelings, good personal development, developed social aspects, empathy, and psychology. Mental health has an influence on behavior (Tengland, 2012), and in the case of students, they are able to adapt to the environment and to the prevailing norms, and able to avoid stress, depression and alcohol abuse, drug abuse and other bad behavior (Hardy et al., 2013). Martin (2012) stated that mental health contributes to learning achievement, perseverance, graduation and overall student success. Students may have difficulties in adjusting themselves. All of these may cause problems for students. Adults in the family, school and community need to understand mental health so that they can be wiser in dealing with, guiding and fostering adolescents. They must help adolescents to get through this phase successfully (Bohnenkamp et al., 2015). Finally, it can be concluded that students' mental health is the existence of genuine harmony between the psychological functions and the ability to adjust according to religious guidance and the ability to actualize the potential.

Literature Review

Mental health

Mace (2007) asserts that health is not only defined as a person's freedom from mental illness but also defined as the ability to adjust, which shows serenity. A behavior is categorized as normal or not normal depending on cultural or community norms. Therefore, normal behavior as evidence of mental health in one community can be deemed not normal for other societies. It means that the state of being normal is not absolute but relative (Townsend & Morgan, 2017). The different response of each individual to the source of stress or problems depends on the individual perception and abilities. In stressful situations, besides having positive emotions, individuals must face negative emotions. Positive emotions can eliminate some of the effects of negative emotions, especially physiological effects. Positive emotion is the energy produced by humans through spiritual meanings that follow stressor events that are able to respond to stressors in a more adaptive way (Kring & Caponigro, 2010; Gonçalves et al., 2015; Szaflarski et al., 2012). Lamborn et al., (2018) said that mental health is a reflection of a good life, including positive feelings that reach an optimal level of life, which can be constructed into three components, namely emotional, social and psychological well-beings.

Gould (2016) identifies four attributes of mental illnesses, namely organic, psychotherapy, sociotherapy and medical. Organic orientation refers to the causes of mental disorders due to physiological abnormalities, while psychotherapy mentions that mental stress is the result of psychological conflict. The association of sociotherapy mentions that mental problems are caused by social and environmental aspects. Lastly, medical model associations are a combination of three orientations namely organic, psychotherapeutic and sociotherapeutic. The fashionable model states that every disease is the result of two environmental factors that integrate into an organism (Gould, 2016). The World Health Organization (WHO) mentions that mental health is an integral part of health and well-being. Mental health is a condition of being healthy physically, psychologically, and socially. It also refers to a condition of being free from mental illness. Determinants of mental health and mental disorders are not only from one's ability to manage thoughts, emotions, behavior, and how a person interacts with others, but also among others social, cultural, economic and political factors (World Health Organization, 2013).

Mental Health Construction

Contemporary literatures agree that the meaning of mental health involves three-dimensional construction, namely emotional, psychological and social. These dimensions have differences in shape and level of well-being. The study of well-beings is divided into two streams, namely equalizing welfare with hedonism or happiness and eudaimonic or human potential that produces positive potential. The hedonic approach relates to the presence of positive emotions, namely by maximizing positive emotions and minimizing negative emotions. Meanwhile, the eudaimonic approach considers optimal psychological function in the life of psychological and social well-being (Petrillo et al., 2015). Previous studies showed that clients from religious backgrounds can increase hedonism (life satisfaction) through eudaimonic well-beings. Religion and spirituality have potential as a source of assistance in counseling, helping to improve adaptability, having goals and optimism in life, and feeling a meaningful life (Yoon et al., 2015; Hefti, 2011).

Religion can affect a person's psychiatric needs by providing a sense of security in the sense of reducing life anxiety (Flannelly & Galek, 2010).The religious component of mental health is associated with holistic mental health therapy, because with a religious approach speeds client's therapy or recovery. Religiosity-based therapy can be more effective than other therapies even with drug therapy, namely by increasing physical and mental health (Chidarikire, 2012; Gonçalves et al., 2015). There are three things that make religion a part of mental health. First, religion with its teachings can guide life and help solve problems. Second, religion has many lessons contained in every religious teaching. Third, religion can be the basis for seeking meaning by feeling the presence of God, feeling the meaning in life,

making a self-identity as a model of goodness in the environment (Behere et al., 2013). As mentioned by the World Health Organization (WHO), mental health is health and well-being, although there are diverse interpretations of the definition of health by including spirituality (Nagase, 2012). Gillam (2018) mentions that being mentally healthy is not only free from mental disorders and mental illness, but there are psychological, physical and social and spiritual well-beings. Spirituality has an important role in mental health. There are a number of notions of mental health, and in principle, these notions make one better than the other, because some of these meanings complement each other so that they become more comprehensive in understanding the meaning of mental health in humans.

Methodology

Research Design

In assessing the mental health instrument, a theoretical and empirical validity test is conducted (Hammarström et al., 2016; McMillan, 2015). Theoretical validation is carried out by 4 experts who work as psychology lecturers and psychometric and evaluation experts and 20 panelists most of whom have completed or are currently in doctoral education in the field of Islamic education and evaluation. They were involved to correct instrument items from the aspect of content or substance, construction and language. The experts consist of 2 people from the field of psychology and 2 people from the measurement of education. The experts conducted a qualitative study, while the panelists reviewed quantitatively (Cecil et al., 2009).

After being refined based on expert suggestions, the instrument was reviewed by panelists to provide a quantitative assessment of the construct by looking at: (1) whether the existing dimensions are components of the construct of a variable, (2) whether the indicators are part of the dimensions, and (3) whether the items in the instrument are the development of the research variable indicators. Panelist validation was carried out using a Likert scale model with five answer choices, and the answers from the panelists were analyzed using the Aiken V index.

$$\text{Aiken V} = \frac{\sum \boldsymbol{n_i}|\boldsymbol{i}-\boldsymbol{lo}|}{[\boldsymbol{N}\ (\boldsymbol{c}-\boldsymbol{1})]}$$

Note: V= validity index from Aiken; $\boldsymbol{n_i}$= number of experts who choose criterion i; r= criterion i; lo= lowest rating; n= the number of all experts; c= number of ratings/criteria

In this study, the validity index was classified as follows: 0.00-0.20 = very low; 0.21-0.40 = low; 0.41-0.60 = sufficient; 0.61-0.80 = high; and, 0.81-1.00 = very high. This theoretical validation is then followed by empirical validation. Empirical validation was carried out on a questionnaire instrument in the form of a 5-point Likert scale, which had five alternative answer choices that were the opposite between positive and negative questions. For positive (favorable/F) questions, the Likert scale classification is: 5 (very suitable), 4 (appropriate), 3 (less suitable), 2 (inappropriate), and 1 (very inappropriate). Conversely, for negative score items (unfavorable/UF), the scores were 1 (very suitable), 2 (suitable), 3 (less suitable), 4 (unsuitable), 5 (very unsuitable) (McMillan, 2015; Linn, 2008). More detailed information of these items can be seen in the Appendix.

Sampling

Empirical validation was done twice to the 16-17-year-old students of Madrasah Aliyah in Jakarta and South Tangerang. The research was conducted in two trials with different sample size in each stage and a test to assess the item validity was conducted in last stage. The number of samples is calculated by referring to Gable and Wolf (2012), that in order to develop a standard instrument, the number of respondents suggested in the try out is 6-10 times the number of items in the instrument. This study took 10 times the number of items as a sampling basis. Taking into account the estimated response rate of 50 percent of the number of questionnaires distributed, the sample size used is:

44 items x 10 x (100/50) = 880, rounded to 850

In the second stage, the number of items was 37 and the estimated response rate was 55 percent, so that the sample size used is:

37 items x 10 x (100/55) = 672, rounded to 650.

In the first stage, the distributed questionnaires were 850 with response rate of 70.24% (597 questionnaires returned), and in second phase, we employed 650 students as sample with response rate of 63.08% (410 questionnaire returned). Third stage was used to analyze the model fitness was conducted to observed fit test of CFA 1 and CFA 2 to assessing 35 valid items.

Data Collection

As explained before, the research was conducted in two trials with different sample size in each stage. The first phase of the trial instrument used a questionnaire containing 3 dimensions, 9 indicators and 44 items. The questionnaire was given to 850 students, 597 of whom returned and filled in the questionnaires. The second phase of the trial included a

questionnaire containing 3 dimensions, 9 indicators and 37 items. The questionnaire was given to 650 students, 410 of whom completed and returned the questionnaires. In the first trial which was responded by 597 students, the analysis of the model evaluation showed that out of 44 items, there were 37 valid items with the criteria having an SLF value ≥0.30 and t-value ≥1.96, and 7 items were invalid. The reliability is calculated with Construct Reliability (CR) = 0.94, and the calculated results show that the items are reliable. The second stage evaluates the structural model, by using the criteria of SLF value ≥0.30, the t-value loading factor ≥1.96, and Construct Reliability (CR)= 0.93> 0.70. In this stage, the number of responses was from 410 students out of 650 questionnaires distributed. The results of 2nd trial showed that from a total of 37 items, there are 2 invalid items, leaving 35 valid items as a basis for the mental health instrument. Furthermore, the third stage of the evaluation of model fitness was conducted to observed fit test of CFA 1 and CFA 2, and the model shows the fit of the mental health instrument model as a whole.

Variable Measurements

The instruments used in this study were adopted from some psychometric measurement and assessment developed by Reynolds et al. (2006), and Cohen et al. (1996). The description of the indicators of the dimensions of a mental health variable is shown in Table 1 (see also, Appendix).

Table 1: Instruments of Mental Health

No	Dimension	Indicators		Points		Σ
				(+)	(-)	
1	Harmony of mental functions	a.	Harmony on feeling	3,4,7,8	1,2,5,6	8
		b.	Harmony in mind	9,12	10,11,13	5
		c.	Harmony in behavior	14, 15	16, 17	4
2	Adaptability	a.	Adjustment to themselves	18, 20, 21	19, 22, 23	6
		b.	Adjustment to others	24, 26, 28	25, 27	5
		c.	Adjustment to the environment	31, 32	29, 30	4
3	Ability to actualize self potential	a.	Learning activities	34, 35	33, 36	4
		b.	Development of interest	39, 40	37, 38	4
		c.	Exercise	41, 43	42, 44	4
Total				22	22	44

Analysis Technique

The testing technique of item validity at the first and second stages was conducted by using Structural Equation Modeling (SEM) analysis. Analysis of the measurement model was done by using Confirmatory Factor Analysis (CFA) with the LISREL 8.80 Full Version program. If the analysis value of Standard Loading Factor (SLF) is smaller than the critical limit of <0.70 or 0.50, the observed verifiability should be removed from the model. In addition to the two critical limits, Igbaria and Baroudi (1993), and Wijayanto (2008) added that if the loading factor is $0.30 \geq x < 0.50$ and t-value ≥1.96, the observed variable can be maintained in the model. Hair Jr. et al. (2014) stated that the valid standard in the CFA is a loading factor ≥ 0.30. This testing technique was used to calculate reliability. Hence, the level of consistency of manifest variables can be determined in measuring the latent construct by using Construct Reliability (CR), and be determined by the acceptance standard of reliability or Construct Reliability (CR)> 0.70 (Kline, 2015).

Results

Theoretical Validity

The expert's theoretical and qualitative validations to mental health instruments of students suggest there are improvements in terms of construction or composition, language and content (Topkaya et al., 2017). The items must be more specific in measuring the indicators of behavior to avoid stereotypical answers. Some items were arranged in favorable directions and some others must be made in unfavorable directions. The experts suggested that 5 items must place the subjects at the beginning of the sentence. The focus of 2 items must be on the respondent's environment and 4 items are suggested to be added. Moreover, 3 items must be substituted, 1 item was requested to be made simpler, 2 items were suggested to be added, 3 items must use standard Indonesian, and 1 word in 3 sentences must be deleted and not be used in the instrument.

The results of the theoretical quantitative validation by the panelists show that 44 items of mental health instruments fit between dimensions, indicators and statement items and are conceptually valid. The panelists validated using the Aiken index, and the results showed that all mental health instrument items were valid, with high (0.61-0.80) and very high (0.81-1.00) calculation results.

Empirical Validity

The item validity test in the first and second stages was conducted using Structural Equation Modeling (SEM) analysis. The analysis of the measurement model was done by using Confirmatory Factor Analysis (CFA) with the Lisrel 8.80 Full Version program. A quality item is known from the loading factor that is on the path coefficient in the measurement model and structural model. Confirmatory Factor Analysis (CFA) by using Lisrel 8.80 Full Version program was used to test the fitness between theoretical models and empirical data. The criteria of Goodness of Fit (GoF) are used as the basis of the measurement model. The cut-off values for GoF in this study are Chi square (p) p-value ≥0.05; Root mean square error of approximation (RMSEA)≤0.08; Expected Cross Validation Index (ECVI)= ECVI Model<ECVI Saturated and Independence model; Anti-image correlation (AIC)= AIC Model <AIC Saturated and Independence Model; CAIC=CAIC Model<CAID Saturated and Independence Model. Other recommended values include Non-normed fit index (NNFI) ≥0.90, Normed fit index (NFI) ≥0.90, Parsimonious normed fit index (PNFI) ≥0.90, Comparative fit index (CFI) ≥0.90, Incremental fit index (IFI) ≥0.90, Relative fit index (RFI) ≥0.90, Goodness of fit index (GFI) ≥0.90, Root mean square Residual (RMR) ≤ 0.08, Adjusted goodness of fit index (AGFI) ≥0.90, and Parsimonious Goodness of Fit Index (PGFI) ≥0.60.

The development of mental health instruments is carried out using first order and second order confirmatory factor analysis, through stages of measurement model evaluation, structural model evaluation, and model fitness testing. The CFA 1 calculation is that mental health instruments contain 37 valid items as observed in Table 2.

Table 2: CFA Model Loading Factors 1 First Test Mental Health Instrument

Factor/Indicator	Manifest	Loading Factor		Error	Verification
		SLF	t value		
Feeling	D 1	0.62	14.52	0.61	Valid
	D 2	0.64	14.90	0.60	Valid
	D 3	0.41	9.24	0.83	Valid
	D 4	0.32	7.15	0.89	Valid
	D 5	0.33	7.36	0.89	Valid
Mind	D 6	0.75	19.60	0.44	Valid
	D 7	0.83	22.31	0.31	Valid
	D 8	0.42	9.83	0.82	Valid
	D 9	0.62	15.30	0.62	Valid
Behavior	D10	0.35	7.53	0.88	Valid
	D11	0.42	9.25	0.82	Valid
	D12	0.59	13.27	0.65	Valid
	D13	0.58	12.94	0.66	Valid
Self-adjustment	D14	0.40	9.45	0.84	Valid
	D15	0.46	11.13	0.79	Valid
	D16	0.33	7.75	0.89	Valid
	D17	0.54	13.16	0.71	Valid
	D18	0.65	16.66	0.57	Valid
	D19	0.58	14.45	0.66	Valid
Adjustment to the Other	D20	0.49	10.91	0.76	Valid
	D21	0.45	9.86	0.80	Valid
	D22	0.72	15.97	0.48	Valid
Environment-adjustment	D23	0.67	16.90	0.55	Valid
	D24	0.72	18.57	0.48	Valid
	D25	0.80	21.24	0.36	Valid
Learning	D26	0.77	19.13	0.41	Valid
	D27	0.42	9.91	0.82	Valid
	D28	0.32	7.41	0.90	Valid
	D29	0.49	11.65	0.76	Valid
Interest	D30	0.60	14.52	0.64	Valid
	D31	0.53	12.59	0.72	Valid
	D32	0.61	15.07	0.62	Valid
	D33	0.61	15.97	0.62	Valid
Exercise	D34	0.68	16.84	0.54	Valid
	D35	0.60	14.33	0.65	Valid
	D36	0.63	15.31	0.60	Valid
	D37	0.76	19.37	0.42	Valid

The results show the accepted standard of reliability, Construct Reliability (CR)> 0.70. Moreover, the result shows that mental health instruments are reliable, with Construct Reliability (CR) of 0.94. Measurement of the second order of Confirmatory Factor Analysis is shown in Figure 1.

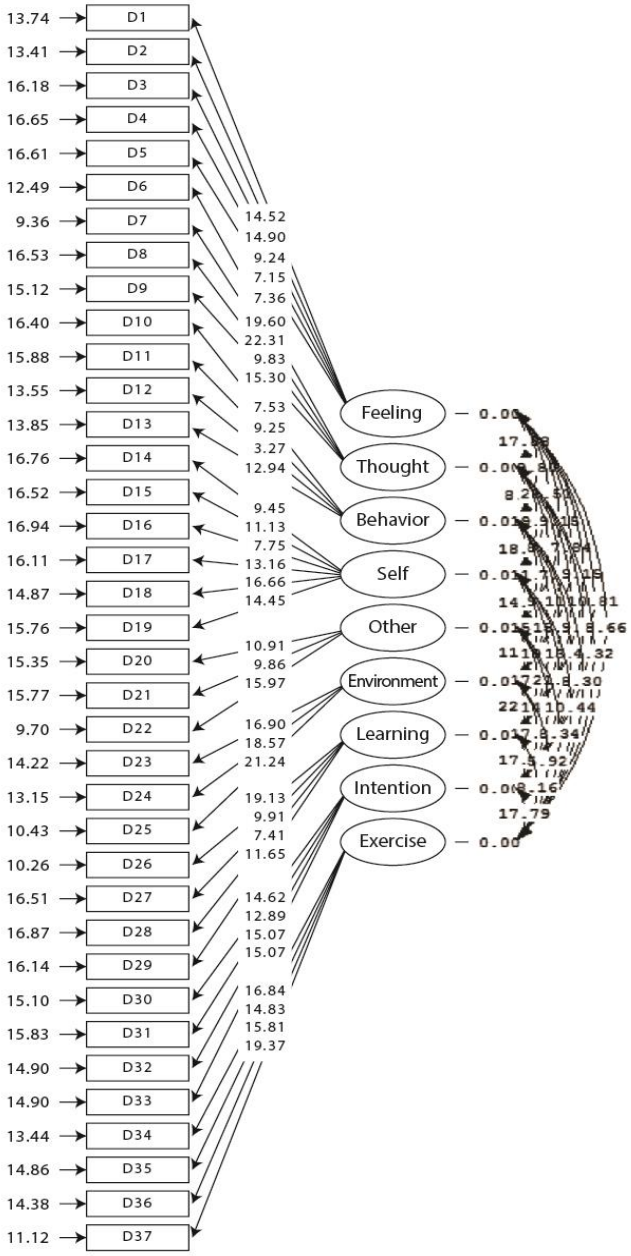

Figure 1: Path Diagram for Second Order of Confirmatory Factor Analysis

Figure 1 presents the measurement of the second order of confirmatory factor analysis with the t-value model, and shows that all items have a t-value >1.96, meaning that all factor loads are significant. Based on Table 3, the model fit test shows that the overall fit of the mental health instrument model.

Table 3.CFA Model-Goodness of Fit Test 1st phase of Mental Health Instrument- 1st Trial

No	Index	Cut-off value	GoF values	Verification
1	Chi square P	*p*-value ≥0.05	*p*-value = 0	Poor Fit
2	Root mean square error of approximation (RMSEA)	≤ 0.08	0.078	Good Fit
3	Expected Cross Validation Index (ECVI)	ECVI Model < ECVI Saturated & Independence model	M = 5 S = 2.36 I = 34.05	Marginal Fit
4	Anti-image correlation (AIC)	AIC Model < AIC Saturated & Independence Model	M = 2977.80 S = 1406 I = 20291.87	Marginal Fit
5	Consistent Akaike Information Index (CAIC)	CAIC Model < CAID Saturated & Independence Model	M = 3570.91 S = 5196.52 I = 20491.37	Good Fit
6	Non-normed fit index (NNFI)	≥ 0.90	0.89	Marginal Fit
7	Normed fit index (NFI)	≥ 0.90	0.88	Marginal Fit
8	Parsimonious normed fit index (PNFI)	≥ 0.90	0.78	Poor Fit
9	Comparative fit index (CFI)	≥0.90	0.91	Good Fit
10	Incremental fit index (IFI)	≥0.90	0.91	Good Fit
11	Relative Fit Index (RFI)	≥0.90	0.86	Marginal Fit
12	Goodness of fit index (GFI)	≥0.90	0.80	Marginal Fit
13	Root mean square Residual (RMR)	≤ 0.08	0.074	Good Fit
14	Adjusted goodness of fit index (AGFI)	≥0.90	0.76	Poor Fit
15	PGFI	≥0.60	0.67	Good Fit

Structural Model Evaluation

The structural model is composed of exogenous latent variables and endogenous variables, which describe the relationship of one variable to another. The validity of the variable construct can be known through the values of the structural model. Figure 2 showed that the observed loading factor of the constructs is valid. The loading value of the feeling is 0.68, mind 0.61. behavior 0.88, self-adjustment 0.97, adjustment to others 0.79, adjustment to the environment 0.70, learning activities 0.88, developing interest 0.88 and exercise 0.52. Thus, it can be concluded that the indicators that measure the latent variables of mental health instruments are valid, with t-value of 1.96 and Standardized Loading Factor (SLF) of 0.30.

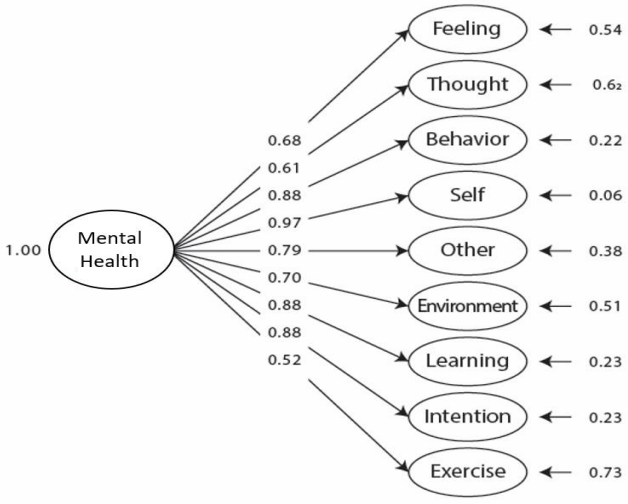

Figure 2: Diagram of the Construct Model of Mental Health Variables

Furthermore, structural model evaluation is done by testing validity and reliability with the calculations shown in Table 4.

Table 4: Validity and Reliability of Mental Health Variables in CFA 2-2nd Test

Indicator	Items	SLF	Error	t-value
Harmony of mental functions	Harmony to feeling	0.68	0.54	9.26
	Harmony in mind	0.61	0.62	10.20
	Harmony in behavior	0.88	0.22	6.70
Adaptability	Self-adjustment	0.97	0.06	8.31
	Adjustment to others	0.79	0.38	8.08
	Adjustment to the environment	0.70	0.51	10.21
Capability to actualize self potential	Learning activities	0.88	0.23	14.77
	Development of interest	0.88	0.23	10.80
	Exercise	0.52	0.73	8.23

Model Fitness Evaluation

The CFA model is used to test the fitness of the model to find out the causal relationship of each latent variable with an indicator or item. The ideal Goodness-of-Fit (GoF) criteria are presented in Table 5.

Table 5: CFA Model Goodness of Fit Model 2- 2nd Test Mental Health Instrument.

No	Size of Feasibility	Cut-off values	GoF values	Verification
1	Chi square (p)	*p*-value ≥0.05	*p*-value = 0	Poor Fit
2	Root mean square error of approximation (RMSEA)	≤ 0.08	0.077	Good Fit
3	Expected Cross Validation Index (ECVI)	ECVI Model < ECVI Saturated and Independence model	M = 5,03 S = 3.08 I = 32.53	Marginal Fit
4	Anti-image correlation (AIC)	AIC Model < AIC Saturated and Independence Model	M = 2059 S = 1260 I = 13305.40	Marginal Fit
5	CAIC	CAIC Model < CAID Saturated and Independence Model	M = 2455.52 S = 4420.18 I = 13480.97	Good Fit
6	Non-normed fit index (NNFI)	≥0.90	0.91	Good Fit
7	Normed fit index (NFI)	≥0.90	0.88	Marginal Fit
8	Parsimonious normed fit index (PNFI)	≥0.90	0.81	Marginal Fit
9	Comparative fit index (CFI)	≥0.90	0.91	Good Fit
10	Incremental fit index (IFI)	≥0.90	0.91	Good Fit
11	Relative fit index (RFI)	≥0.90	0.87	Marginal Fit
12	Goodness of fit index (GFI)	≥0.90	0.79	Poor Fit
13	Root mean square Residual (RMR)	≤ 0.08	0.074	Good Fit
14	Adjusted goodness of fit index (AGFI)	≥0.90	0.76	Poor Fit
15	Parsimonious Goodness of Fit Index (PGFI)	≥0.60	0.69	Good Fit

Based on Table 5, it can be seen that there are 3 indices with poor fit, 5 indices with good fit, and 7 indices with good fit. From the results of the model fit test, it can be seen that the fitness of the overall mental health instrument model is good.

The theoretical and empirical validation processes that have been carried out on mental health instruments from the validation results show that there are 7 invalid items from the first trial, and 2 invalid items in the second trial. This reduces the number of items from 44 items to 35 items. Furthermore, 35 valid items were compiled into a final set of mental health instruments, which form the basis of the questionnaire structure (Appendix). The mental health instrument grid is shown in Table 6.

Table 6: Mental Health Instruments after Validation

No	Indicator	Items		Item Number		Σ
				(+)	(-)	
1	Harmony of mental functions	a.	Harmony in feeling		1, 2, 3, 4	4
		b.	Harmony in mind	7	5, 6, 8	4
		c.	Harmony in behavior	9	10, 11	3
2	Adaptability	d.	Adjustment to themselves	12, 14, 15	13, 16,17	6
		e.	Adjustment to others	19	18, 20	3
		f.	Adjustment to the environment	22, 23	21	3
3	Capability to actualize self-potential	g.	Learning activities	25, 26	24, 27	4
		h.	Development of interest	30, 31	28, 29	4
		i.	Exercise	32, 34	33, 35	4
Total						**35**

Lastly, the mental health assessment is conducted regarding the overall result regarding the general description of the mental health condition of students from all stages of the 35 questionnaire items. The basis for the assessment for measuring the student mental health involved in this study was obtained from the highest and lowest scores for each trial, and were divided into 3 classifications (good, moderate, and less). The highest and lowest scores in the first stage were 181 and 75, respectively. Meanwhile, in the second stage, the highest score was 175, and the lowest score was 91 (Table 7). The results of the research in the first and second trials showed that the majority of mental health students at Madrasah Aliyah (MA) were in the adequate mental health category. The results of the first trial of 597 students, there were 416 students or 69.68% in the moderate mental health category, 140 students or 23.45% in the good mental health category, and 41 students or 6, 87% in the health category mentally less. Furthermore, the results of the 2nd trial out of 410 students, there were 252 students or 61.46% in the moderate mental health category, 133 students or 32.44% in the mental health category, and 25 students or 6, 10% in the category good mental health. Based on these data it can be concluded that most of Madrasah Aliyah (MA) students have adequate mental health. This situation can be caused because students of Madrasah Aliyah (MA) are in their teens. Adolescence is a period of transition from childhood and preparation for adulthood. Along with this period, there was rapid physical and psychological development because adolescents were not yet able to master and function properly their physical and psychological functions.

Table 7. Results of 1st and 2nd Try Out (Classification of 3 groups)

Score	Category	Freq.	Percentage
1st try out:			
146 - 181	Good	140	23.45
111 - 145	Moderate	416	69.68
75 - 110	Less	41	6.87
2nd try out:			
148 - 175	Good	25	6.10
120 - 147	Moderate	252	61.46
91 - 119	Less	133	32.44

The assessment generally denoted that emotional stability that has not been formed properly in adolescence is likely able to affect psychological process. Here, the role of the environment is needed which can help as a preventive and constructive measures to improve adolescent mental health, and to minimize disturbances in their mental health. A good environment that adolescents need is a family, school and community environment that can meet their psychological needs. The specific needs of these environments include providing a social atmosphere, healthy and safe conditions, emotional and behavioral education, and counseling on religious values. The role of adults is also needed by understanding and responding well to the situation of adolescents who are experiencing a sensitive period from their psychological development.

Discussion

This study analyzes the stages of the development and validation of standard instruments as an analytical tool to measure the mental health of school age adolescents. In this study, empirical validation or testing of the instruments was carried out in the field. Test the validity of the items empirically at the first and second stages using Structural Equation Modeling (SEM) analysis. The measurement model analysis was performed using the Confirmatory Factor Analysis (CFA), namely CFA 1 and CFA 2 with the Lisrel 8.80 Full Version program. Measurement of items is carried out

by paying attention to the concept of a measured variable construct, namely the mental health variable. The reference for the measurement and utilization of good or quality instruments in general is that it has requirements, namely valid and reliable (Miller et al., 2009; Ronald Jay Cohen & Swerdlik, 2010). The findings of this study confirm the validity of the items of adolescent mental health, particularly students of Madrasah Aliyah in Jakarta and South Tangerang. This is in accordance with Cecil et al. (2009) who state that a valid instrument is to have fitness in performing its measuring function, and the reliable instruments are able to measure results with a stable or consistent measuring score. Furthermore, the use of panelists' opinions in this study is used to provide an opinion regarding the fitness of items to the context of adolescent mental health, and to modify them. This is in accordance with Kubiszyn and Borich (2010), who found that the validity and reliability of an instrument are fulfilled, indicating that the instrument is standardized. Standardized instruments are compiled through a rational validation process by experts and panels which are then refined according to the suggestions of experts and panels. After that, the empirical validation process is carried out or tested, calculated, analyzed and assembled into a set of valid and reliable instruments.

In this study, empirical validation is done twice for students of Madrasah Aliyah in South Jakarta and South Tangerang. The item validity test technique in the first and second stage trials uses Structural Equation Modeling (SEM) analysis. Analysis of the measurement model is done by using the Analysis Factor Confirmatory (CFA) with the LISREL 8.80 Full Version program. A quality item is known from the loading factors that exist in the path coefficients in the measurement model and structural models. Item criteria are said to be valid if the Standard Loading Factor (SLF) is ≥0.30 with a t-value ≥1.96. Furthermore, to calculate reliability uses SEM analysis. It is the evaluation of reliability measurement with Construct Reliability so that the level of consistency of manifest variables can be identified in measuring the latent construct. The standard used for the acceptance for reliability is with the value of construct reliability (CR)> 0.70. The results confirmed that mental health instruments obtained from observations of the fitness of the CFA 1 and CFA 2 models in the second trial showed the fitness of the overall mental health instrument model, having a good model fit, i.e. having the same 7 GoF indices indicating good fit on size feasibility of RMSEA, CAIC, NNFI, CFI, IFI, RMR, and PGFI. The results of this study confirm Hudson et al. (2020) and Arici-Ozcan et al. (2019) who found a relationship between psychological stress experienced by adolescents and the influence of cognitive factors to strengthen the mental health of school age adolescents. In the context of mental health development, there have been several previous studies that analyzed the relationship between validation and measurement with educational performance in general. In the context of secondary schools in Indonesia, this study highlights aspects of measurement validation, which were also expressed by Abubakar et al. (2015). Furthermore, Tran et al. (2019) analyzed the anxiety disorders among adolescent in Indonesia. The findings of this study are also in line with Garey et al. (2020) who evaluated role of religious orientation on the mental health of adolescence in Indonesia.

Conclusion

The development of mental health instruments was carried out through the process of theoretical validation and empirical validation. The theoretical validation was done by experts and panelists that examine items in terms of construction, content and language, as well as analyzing whether representative points represent indicators. Indicators are precise descriptions of dimensions. Dimensions are the exact description of the operational definition. The concept definition of the variable represents the construct of a mental health variable, which is the synthesis of theories. The results of a review of 4 qualitative experts showed that they generally stated that they were good. There were only a few suggestions from experts on a small portion of mental health instrument items to be revised in terms of construction, content and language. This, it could be concluded in principle that there were no meaningful changes, after the points were fixed according to expert advice, then theoretical validation was carried out by 20 panelists quantitatively using the Aiken index. The calculated results showed that all instrument items were valid, with high interpretation of the calculated results (0.61-0.80) and very high (0. 81-1.00). The results showed the model fitness evaluation, the fit test observations the CFA 1 and CFA 2 models show the same results. In conclusion, the fitness of the overall mental health instrument model has a good model fit.

As recommendations, the findings of this study focus on the validity and reliability of adolescent mental health in Indonesia. The findings indicate practical and theoretical recommendations that need to be made in relation to the findings. This study theoretically contributes to measuring items that are valid for adolescent mental health with items such as feeling, mind, behavior, self, relationships with other people, environment, learn, interest and exercise. Practically, this study underlines the importance of the elements of harmony of mental functions, adaptability and the ability to actualize self-potential in the formation of adolescent mental health.

Finally, although this study is insightful in providing a measure of construct validation and reliability that can be used for further research on adolescent health, several limitations of this study have limited its applicability. Firstly, since this study was conducted in Islamic high school, or Madrasah Aliyah (MA), the limitation is that the generalization of this study is limited to adolescent in Islamic high school student, age from 15 to 18 years old. Secondly, this study was conducted by focusing on the relationship between mental health variables and their various constructs. Although this study provides clear and measurable empirical evidence regarding the validity and reliability of the various items used, the lack of testing between items and between variables makes it less likely to adopt them in direct testing. Thirdly, this

study did not classify the demographic characteristics of the sample as the basis for cross-sectional testing. Furthermore, this test is limited to adolescents at the high school level, in formal religious education institutions, in Indonesia. Future research is expected to broaden the baseline of measurement for adolescents over a wider age range, and use the demographic characteristics of respondents to analyze differences in mental health across demographic elements. Lastly, further studies also need to calculate the external validity in which the instrument is utilized to mentally healthy and mentally unhealthy adolescents to see the discriminating power of the instrument.

Acknowledgment

We thank the Division Head of Islamic Education School of the Provincial Branch of Ministry of Religion of Jakarta, Headmasters of Islamic Senior High School in South Jakarta (MAN 4, MAN 13, MAN 19, MAN 7, MAN 11) and Headmaster of Islamic Senior High School Pembangunan of Syarif Hidayatullah State Islamic University Jakarta and Headmaster of An Najah Islamic Senior High School, for the permission that give opportunity for the authors to conduct research. We also thank all the students of the Islamic Senior High School who have participated as respondents in our research. We would also like to thank the Faculty of Islamic Education and Teacher Training (FITK) of UIN Syarif Hidayatullah Jakarta and Jakarta State University for their generous support for this research. Our sincere thanks also go to the anonymous reviewers who have provided constructive suggestions on this paper.

References

Abubakar, A., Van de Vijver, F. J., Suryani, A. O., Handayani, P., & Pandia, W. S. (2015). Perceptions of parenting styles and their associations with mental health and life satisfaction among urban Indonesian adolescents. *Journal of Child and Family Studies*, *24*(9), 2680-2692.

Arici-Ozcan, N., Cekici, F., & Arslan, R. (2019). The relationship between resilience and distress tolerance in college students: The mediator role of cognitive flexibility and difficulties in emotion regulation. *International Journal of Educational Methodology, 5*(4), 525-533. htpps://doi.org/ 10.12973/ijem.5.4.525

Behere, P. B., Das, A., Yadav, R., & Behere, A. P. (2013). Religion and mental health. *Indian journal of psychiatry*, *55*(Suppl 2), S187–S194.

Bohnenkamp, J. H., Stephan, S. H., & Bobo, N. (2015). Supporting student mental health: The role of the school nurse in coordinated school mental health care. *Psychology in the Schools*, *52*(7), 714-727.

Cecil, R., Ronald, B. R., & Victor, W. (2009). *Measurement and Assessment in Education.* Pearson Education.

Chidarikire, S. (2012). Spirituality: The neglected dimension of holistic mental health care. *Advances in Mental Health*, *10*(3), 298-302.

Clauss-Ehlers, C. S., Serpell, Z. N., & Weist, M. D. (2013). Introduction: Making the case for culturally responsive school mental health. In *Handbook of Culturally Responsive School Mental Health* (pp. 3-15). Springer.

Cohen, R. J., & Swerdlik, M. E. (2010). *Psychological Testing and Assessment*. Tata McGraw Hill Education.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.

de Almeida Vieira, A. P. T., & Fernandes, A. B. (2016). Cultural competence in mental health nursing: validity and internal consistency of the Portuguese version of the multicultural mental health awareness scale—MMHAS. *BMC psychiatry, 16*(1), 1-8.

Eryılmaz, A., & Sapsağlam, Ö. (2018). Development of subjective well-being increasing strategies scale for children (1-5 ages) their mothers use. *Journal of Education and Training Studies*, *6*(6), 77-83.

Flannelly, K. J., & Galek, K. (2010). Religion, evolution, and mental health: Attachment theory and ETAS theory. *Journal of Religion and Health*, *49*(3), 337-350.

Gable, R. K., & Wolf, M. B. (2012). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings* (Vol. 36). Springer Science & Business Media.

Garey, E., Siregar, J. R., Hood Jr, R. W., Agustiani, H., & Setiono, K. (2016). Development and validation of Religious Attribution Scale: In association with religiosity and meaning in life among economically disadvantaged adolescents in Indonesia. *Mental Health, Religion & Culture*, *19*(8), 818-832.

Gillam, T. (2018). *Creativity, Wellbeing and Mental Health Practice*. Palgrave Pivot.

Gonçalves, J. P., Lucchetti, G., Menezes, P. R., & Vallada, H. (2015). Religious and spiritual interventions in mental health care: a systematic review and meta-analysis of randomized controlled clinical trials. *Psychological Medicine*, *45*(14), 2937-2949.

Gonçalves, T. E. D., Zimmermann, G. S., Figueiredo, L. C., Souza, M. D. C., da Cruz, D. F., Bastos, M. F., da Silva, H. D. P., & Duarte, P. M. (2015). Local and serum levels of adipokines in patients with obesity after periodontal therapy: one-year follow-up. *Journal of Clinical Periodontology*, *42*(5), 431-439.

Gould, N. (2016). *Mental health social work in context*. Routledge.

Hair Jr. J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. Pearson Education.

Hammarström, A., Westerlund, H., Kirves, K., Nygren, K., Virtanen, P., & Hägglöf, B. (2016). Addressing challenges of validity and internal consistency of mental health indices in a 27-year longitudinal cohort study–the Northern Swedish Cohort study. *BMC Medical Research Methodology*, *16*(1), 1-9.

Hardy, S. A., Francis, S. W., Zamboanga, B. L., Kim, S. Y., Anderson, S. G., & Forthun, L. F. (2013). The roles of identity formation and moral identity in college student mental health, health-risk behaviors, and psychological well-being. *Journal of Clinical Psychology*, *69*(4), 364-382.

Hefti, R. (2011). Integrating religion and spirituality into mental health care, psychiatry and psychotherapy. *Religions*, *2*(4), 611-627.

Hudson, P. J., Lai, B., & Short, M. B. (2020). Cumulative risk of psychological distress in college students effected by Hurricane Harvey. *European Journal of Psychology and Educational Research, 3*(2), 101-109.

Igbaria, M., & Baroudi, J. J. (1993). A short-form measure of career orientations: A psychometric evaluation. *Journal of Management Information Systems, 10*(2), 131-154.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kring, A. M., & Caponigro, J. M. (2010). Emotion in schizophrenia: where feeling meets thinking. *Current Directions in Psychological Science*, *19*(4), 255-259.

Kubiszyn, T., & Borich, G. (1987). *Educational testing and measurement*. Scott, Foresman.

Lamborn, P., Cramer, K. M., & Riberdy, A. (2018). The structural validity and measurement invariance of the mental health continuum–short form (MHC-SF) in a large Canadian sample. *Journal of Well-Being Assessment*, *2*(1), 1-19.

Linn, R. L. (2008). *Measurement and assessment in teaching*. Pearson Education India.

Liu, C., Yanling, L., Guo, C., & Lan, H. (2013). Preliminary exploration of the mental health education competency survey of primary and middle school head teachers. *Journal of Education and Training Studies*, *2*(1), 73-80.

Mace, C. (2007). *Mindfulness and mental health: Therapy, theory and science*. Routledge.

Martin, D. K. (2012). *Measuring the impact of a residential learning community on the mental health and well-being of art students in higher education* [Unpublished doctoral dissertation]. Loyola University Chicago.

Maslow, A. H. (1981). *Motivation and personality*. Harper & Row.

McMillan, J. H. (2015). *Fundamentals of educational research*. Pearson.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching*. Pearson Education.

Nagase, M. (2012). Does a Multi-Dimensional Concept of Health Include Spirituality? Analysis of Japan Health Science Council's Discussionson WHO's 'Definition of Health' (1998). *International Journal of Applied Sociology, 2*(6), 71-77.

Omoniyi, M. B. I. (2016). Unemployment and underemployment as indices of Nigerian youths' mental health and the place of agricultural revolution as a Panacea: Implications for counselling. *Journal of Education and Practice*, *7*(10), 80-88.

Petrillo, G., Capone, V., Caso, D., & Keyes, C. L. (2015). The Mental Health Continuum–Short Form (MHC–SF) as a measure of well-being in the Italian context. *Social Indicators Research*, *121*(1), 291-312.

Reynolds, C. R., Livingston, R. B., & Willson, V. (2009). Measurement and assessment in education. Merrill.

Reynolds, C. R., Livingston, R. B., & Willson, V.L. (2006). *Measurement and Assessment in Education*. Pearson Education.

Szaflarski, M., Kudel, I., Cotton, S., Leonard, A. C., Tsevat, J., & Ritchey, P. N. (2012). Multidimensional assessment of spirituality/religion in patients with HIV: conceptual framework and empirical refinement. *Journal of Religion and Health*, *51*(4), 1239-1260.

Tengland, P. A. (2012). Health and morality: two conceptually distinct categories? *Health Care Analysis*, *20*(1), 66-83.

Topkaya, N., Sahin, E., & Meydan, B. (2017). The development, validity, and reliability of the Barriers to Seeking Psychological Help Scale for college students. *International Journal of Higher Education*, *6*(1), 48-62.

Townsend, M. C., & Morgan, K. I. (2017). *Psychiatric mental health nursing: Concepts of care in evidence-based practice*. FA Davis.

Tran, T. D., Kaligis, F., Wiguna, T., Willenberg, L., Nguyen, H. T. M., Luchters, S., Azzopardi, P., & Fisher, J. (2019). Screening for depressive and anxiety disorders among adolescents in Indonesia: Formal validation of the centre for epidemiologic studies depression scale–revised and the Kessler psychological distress scale. *Journal of Affective Disorders*, *246*, 189-194. https://doi.org/10.1016/j.jad.2018.12.042

Wijayanto, S. H. (2008). *Structure equation modelling, konsep and tutorial dengan LISREL 8.80* [Structural equation modeling, concepts and tutorials with LISREL 8.80]. Graha Ilmu.

World Health Organization. (2013). *World Health Organization mental health action plan 2013–2020*. https://www.who.int/publications/i/item/9789241506021.

Yoon, E., Chang, C. C. T., Clawson, A., Knoll, M., Aydin, F., Barsigian, L., & Hughes, K. (2015). Religiousness, spirituality, and eudaimonic and hedonic well-being. *Counselling Psychology Quarterly*, *28*(2), 132-149. https://doi.org/10.1080/09515070.2014.968528

Appendix

Items used for measuring mental health with the explanation of stage by stage validation

No	Indicator	F/UF*	Items	Validity	Questionnaire Numbering
1	Feeling	UF	For no apparent reason I get anxious easily	Valid	1
2		UF	I hesitate in making a decision	Valid	2
3		F	I am not a person who panics easily over unfinished tasks	Valid	3
4		F	Feelings of happiness colored myself for the achievements I had achieved	Invalid	*omitted*
5		UF	I was pressured to face multiple assignments	Invalid	*omitted*
6		UF	I am suspicious in social situations	Valid	4
7		F	I am not a person who is easily offended in relationships	Invalid	*omitted*
8		F	I can accept the failures that occur in life	Invalid	*omitted*
9	Mind	F	My study concentration is not distracted by the problems I am facing	Invalid	*omitted*
10		UF	There are problems in life, making my thinking ability decrease	Valid	5
11		UF	Many problems in life make me tired to think	Valid	6
12		F	I have a good memory, not bothered by the problems I am facing	Valid	7
13		UF	I find it difficult to think clearly	Valid	8
14	Behavior	F	I prefer traveling with family or friends rather than going alone	Invalid	*omitted*
15		F	Friends or other people feel happy with my presence	Valid	9
16		UF	I speak and act without thinking	Valid	10
17		UF	I prefer to be alone than to meet and hang out with friends	Valid	11
18	Self	F	I accept my shortcomings gracefully	Valid	12
19		UF	I don't know my own strengths and weaknesses	Valid	13
20		F	I learned a lot from previous experiences	Valid	14
21		F	With my abilities, I believe I can achieve my goals	Valid	15
22		UF	My shortcomings make me less confident	Valid	16
23		UF	I wish I could be born again in a situation that is not like this	Valid	17
24	Other	F	I can feel what other people feel	Invalid	*omitted*
25		UF	The shortcomings that exist in other people prevent me from being able to cooperate	Valid	18
26		F	Religious differences do not prevent me from making friends	Valid	19
27		UF	I feel uncomfortable hanging out with people of different ethnicities or regions	Valid	20
28		F	I am open to other people's opinions	Invalid	*omitted*
29	Environment	UF	I broke the rules at school	Valid	21
30		UF	I find it difficult to adapt to a new environment	Invalid	*omitted*
31		F	I obey the rules or norms that exist in society	Valid	22
32		F	I obey the rules that exist in the family environment	Valid	23
33	Learning	UF	I neglected lessons at school	Valid	24
34		F		Valid	25
35		F	I try to deepen the lessons received at school	Valid	26
36		UF	I enjoy observing what is and is happening in the environment	Valid	27
37	Interests	UF	I feel that it is not important to participate with friends in discussing learning activities at school	Valid	28
38		UF	The extracurricular activities I chose didn't fit my interests	Valid	29
39		F	My hobby is not very useful	Valid	30
40		F	I have an interest that supports learning achievement at school	Valid	31
41	Exercise	F	I regularly exercise	Valid	32
42		UF	In my daily life I don't have time to exercise	Valid	33
43		F	I am a person who likes to exercise	Valid	34
44		UF	I avoid exercise	Valid	35
Total Valid/Invalid				35/9	35

*Favorable (F)/unfavorable (UF)