



The Instrument Construction for Students' Religious Moderation: Psychometric Analysis

Al Halik^{*)}¹, Aisyah Khumairo², Umi Yawisah³

^{1 2 3} Department of Islamic Guidance and Counseling, Institut Agama Islam Negeri (IAIN) Metro, Lampung, Indonesia

*) Corresponding author,  e-mail: alhalik@metrouniv.ac.id

Received: 21 October 2022	Accepted: 06 December 2022	Published: 31 December 2022
------------------------------	-------------------------------	--------------------------------

Abstract

Religious moderation is becoming a prevalent focus to achieve social harmony and peace, preventing extremism, intolerance, and acts of violence. The current study aims to investigate the psychometric properties of an instrument measuring students' religious moderation attitudes. The research method used is the quantitative construction of measuring instruments. The scale development indicator uses the four dimensions of religious moderation developed by the Indonesian Ministry of Religion. The scale development procedure involved 10 experts consisting of academics and employees of the Metro City Ministry of Religion. 135 Respondents for the instrument trials were drawn using a random sampling technique, taking into account students of IAIN Metro in semesters 3, 5, and 7 who were still active. The results show that based on the consistency value of experts, which is known through the interquartile range (Q), there are 60 items out of 78 declared significant. The content validation category based on Aiken's V count is 0.693, which is interpreted as a moderate category. The items have been validated by experts and fulfilled the discriminatory power criteria according to the psychometric analysis prerequisites. Based on the validity test results, using the baseline correlation on 60 items, 36 items fulfilled the validity level. Reliability testing shows a coefficient of 0.800 (reliable) means that this scale is consistent. These findings are expected to benefit the measurement of students' religious moderation attitudes and serve as a need assessment in the establishment of regulations and policies in higher education.

Keywords: Thurstone Scale, Psychometry, Religious Moderation

Abstrak

Moderasi beragama menjadi topik kebijakan pemerintah Indonesia dalam menangkal radikalisme. Artikel ini mengkaji pengembangan alat ukur sikap moderasi beragama mahasiswa yang diuji dari perspektif psikometri. Metode penelitian yang digunakan kuantitatif konstruksi alat ukur. Pengembangan skala melibatkan panelis (ahli) yang berjumlah 10 orang yang terdiri dari akademisi dan pegawai Kementerian Agama Kota Metro. Responden ujicoba instrumen yaitu mahasiswa IAIN Metro yang masih aktif, ditarik dengan teknik purposive sampling sehingga diperoleh 135 orang mahasiswa. Hasil menunjukkan nilai konsistensi para ahli dapat diketahui melalui nilai interkuartil (Q) yaitu terdapat 60 butir dari 78 butir yang dinyatakan signifikan karena memiliki nilai kecil dari 4 (<4). Kategori validasi isi berdasarkan hitungan V Aiken sebesar 0,693 yang dapat diinterpretasikan sebagai kategori sedang. Butir skala telah tervalidasi para ahli, terpenuhi daya beda sesuai prasyarat analisis psikometri. Berdasarkan hasil uji validitas menggunakan korelasi baserial pada 60 butir maka didapatkan 36 butir yang memenuhi tingkat validitas. Pengujian reliabilitas menunjukkan kekonsistenan dan tinggi. Temuan ini diharapkan dapat bermanfaat untuk mengukur sikap moderasi beragama mahasiswa dan menjadi analisis kebutuhan serta menjadi acuan dalam pembentukan regulasi dan kebijakan di perguruan tinggi.

Kata Kunci: Skala Thurstone, Psikometri, Moderasi Beragama



This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author.

INTRODUCTION

The study of religious moderation has received a lot of attention since Indonesia has a pluralistic multicultural heritage that includes various ethnicities, cultures, languages, customs, beliefs, and religions. Positively, this diversity can be one of the country's cultural treasures. However, this also might instigate discord and disintegration. The nation was then united on the basis of Pancasila. However, the country still faces difficulties in trying to integrate Pancasila values. Apart from having the largest Muslim majority society and the Pancasila democratic system, Indonesia is a country that has successfully transitioned to democracy. However, Indonesia also has the potential to become a dangerous precedent for the growth of radicalism and religious intolerance in the context of democracy (Hadiz, 2016). Azyumardi Azra believed that the emergence of radical groups in Indonesia was caused by the euphoria of democracy, thus providing space for extremist groups to express their ideas and activities (Azra, 2003).

Prevention of this is by being wasathiyah, because it will lead to an inclusive, accommodative and tolerant pattern of Muslims towards people of other religions (Kholis et al., 2020). This is important considering Indonesia is a pluralistic country, plural and diverse. Without a wasathiyah attitude, Indonesia would certainly be full of separations and disputes which could destroy the order of life. Furthermore, moderation in religion with the path of Wasathiyah Islam will encourage Muslims to love peace, prevent takfirism (ideas that easily make others unbelievers). According to Azra, by being wasathiyah, Indonesian Muslims will be able to accept four fundamental of the state structure, the NKRI (Negara Kesatuan

Republik Indonesia), Pancasila, UUD 1945 and Bhinneka Tunggal Ika (Azra, 2017).

Religious moderation is an essential and significant as a framework for managing religious life and maintaining unity and oneness (Abdullah et al., 2020). The practical choice for Indonesia, a country with many different religions, is religious moderation. All children of the nation, especially students who will become the nation's next generation, must maintain a balance between national pride and freedom of religion (Rijal et al., 2022). Justifiably, the concept of moderation is the policies foundation to combat extreme religious narratives (Aslam & Gunaratna, 2019), without exception in higher education institutions.

Higher education is an important component in creating and popularizing the idea of religious moderation and in this case is the Islamic Religious College or PTKI, because PTKI is under the Ministry of Religion which develops the idea of religious moderation. PTKI has become a center for Islamic studies because Islam is the majority religion in Indonesia, and PTKI as a forum for producing Muslim scholars, and Islamic boarding schools as a strategy for higher education management to transmit Islamic values (Nasir & Rijal, 2021). The Directorate of Islamic Education of the Ministry of Religion issued circular number B2 3663.1/Dj.I/BA.02/10/2019 dated 29 October 2019 about the House of Religious Moderation or RMB. RMB is the implementing agency for strengthening religious moderation in the PTKI environment which is assigned to be a center for education, mentoring, complaints and strengthening the discourse of the religious moderation movement.

One of the basic elements of moderate Islam in Indonesia is Islamic Higher Education. IAIN alumni with a moderate frame of mind in the development philosophy and modernization ideology that dominated in the 1980s (Jabali & Jamhari, 2002). However, research findings in recent years regarding violent extremism in tertiary institutions also occur in Islamic Religious Colleges (Abdi, n.d). The Wahid Institute survey of 150 million adults stated that issues of intolerance and radicalism have recently become a trend that tends to increase (Antara, 2020). Several studies have shown that not only adults, young people already have radical attitudes, they have even entered several campuses in Indonesia. Survei Mata Air Fondation and Alvara The Research Center and Minister of Defense Ryamizard said 23.4% of Indonesian students were exposed to radicalism (Ali & Purwandi, 2016). LIPI researchers stated that the growth and development of radical understanding in educated communities such as campuses was through the recruitment of cadres by utilizing student organization discussion forums (Hasanuddin, n.d).

Indicators of violent extremism namely; there are transnational ideologies and ideas that seek to establish an Islamic state (dawlah Islamiyah) or khilafah, as well as takfiriism-disbelief for those who adhere to the same religion but follow different religious understandings and practices as well as, the prohibition of tolerance and good relations with adherents of other religions (Schmid, 2015), must be supported by the notion of religious moderation as a powerful discourse. This study confirms the findings of previous research showing the existence of violent extremism in Islamic religious colleges. Therefore, it is necessary to take

action in tertiary institutions, one of which is by maximizing the role of the House of Moderation of Religion (PPIM UIN Jakarta, 2021).

The phenomenon mentioned above is worrying and gives an image like an iceberg, a big problem but imperceptible and needs serious attention, especially in the young generation in college. Therefore, higher education should be able to provide religious moderation education (Murtafallo, 2020). A national standard is needed to represent Indonesia's level of religious moderation when determining the target index to achieve it. It is important to evaluate religious moderation as a means to see the nation from a religious and spiritual perspective. The anti-extreme attitude towards religious ideology must also be a product of religious moderation. In addition, religious moderation needs to go beyond formal practices such as ritual and non-ritual devotion (Barker, 2022).

The first step in preventive efforts is to explore and measure the extent of religious moderation. A valid and reliable measurement should be used for indexing (Furr & Bacharach, 2008; Zufriani et al., 2022) and articulating the goal of achieving religious moderation (instrument). Thus, there is need of measurement to assess the high and low attitudes of individual religious moderation.

The aim of current research is to develop an instrument attitude measurement tool which created in the form of the Thurstone Equal Appearing Interval scale to measure students' attitudes of religious moderation. As Thurstone scaling can convert subjective individual ranking order data or comparative preference data to single-group composite interval scales (Krabbe, 2008).

The Thurstone scale is synonymous with involving experts or judgment to assess each statement item and to ask for an assessment of the location of the favorable level of the statement on a psychological number that can predict the results being measured (Guffey et al., 2007) and not for approval (Azwar, 2013). In addition, the Thurstone scale has the advantage of avoiding social desirability (dishonest responses) and faking (deliberately changing self-characteristics) (Chernyshenko et al., 2009; McDonald, 2013). Because this scale provides an opportunity for the subject to choose only one statement that is appropriate and not appropriate from the several statements it provides (McDonald, 2013).

Instruments to measure religious moderation have indeed been made. Some only measure aspects of religiosity and community tolerance (Abu-Raiya & Hill, 2014; Amir, 2021) and focusing on the scale of tolerance (Meiza, 2018). Munir and Herianto (2020) developed a questionnaire on understanding moderation in religion students which not using an attitude scale. Zufriani et al (2022) in developing the instrument, getting advice from expert judgment to use a scale with agree and disagree options, because the use of options such as (strongly agree, agree, neutral, disagree, strongly disagree) causes informants to be more inclined to choose a neutral option to avoid problems.

Pratama, developed an instrument for measuring high school student religious moderation, which in developing the instrument follows the steps developed by Kyriazos and Stalikas (Pratama, 2020). Whereas in this study the development of the instrument used the stages developed by Azwar (2013). The development follows the procedure of psychometric characteristics with validation results and

an acceptable level of reliability. Insofar as the advantages are possessed, the disadvantages found in the development of the Primary scale are that they have not yet carried out a Vaiken validation test. Instruments that require expert judgment are advised to analyze the calculation of coefficient V or Aiken Validation (Aiken, 1985). One of the good psychometrics, marked by statement items that can distinguish individuals who have high scores from individuals who have low scores. The instrument that was developed by Pratama, has not yet found Aiken's V test and also grain discriminant analysis.

When evaluating test results, a psychometric approach is the methodology used to address measurement issues. According Kaplan and Saccuzo (2017), items in a good test is a good test tool. A number of factors, including validity, reliability, item difficulty index, item discrimination index, and distractor effectiveness, can be analyzed to determine the quality of an item.

Therefore, the purpose of this study was to develop an instrument on the Thurstone Equal Appearing Interval scale to measure students' attitudes of religious moderation and conducting psychometric characteristic analysis in the form of validity, discriminating power and reliability tests. This is deemed necessary, because there are many indications of intolerant students, radicalism and fundamentalism. It is even suspected that students who are intolerant have occurred since when they were still students (Murtadlo, 2020). The iceberg phenomenon found from several studies on the development of radicalism, intolerance and violence in students, requires careful efforts designing comprehensive map of religious moderation, so that it can be used as a reference for preparing policies and

preventing radicalism, intolerance and violence in universities.

METHODS

The research approach used is quantitative descriptive measuring instrument construction. The indicators of religious moderation follow the indicators developed by the Indonesian Ministry of Religion which consist of national commitment, tolerance, non-violence and accommodation to local culture. The four indicators are then developed using the Thurston scale. Followed by expert validation tests, then tested on respondents in the field, carried out the scaling process, and continued with psychometric characteristic analysis. The expert validation completed by 10 experts consisting of lecturers from IAIN Metro, the Metro City Indonesian Ulema Council (MUI), and Metro City Ministry of Religion employees with qualified background consists of the fields of Education, Islamic Studies, Islamic Law, Islamic Politics, Languages, Religious Education. The instrument trial subjects were active students of IAIN Metro. The collection was carried out using a random sampling technique, namely students in semesters 3, 5 and 7 who are still actively participating in lectures. The data collected from 135 students spread across four faculties.

Data analysis in this study is the first consideration in selecting items based on expert judgment by calculating the median value and interquartile value (Q) (Lisnawati, 2011). Aiken suggests using the Aiken validity calculation or the V coefficient (Aiken, 1985). The validity of the item uses the content validity test from V

Aiken. The items that have been prepared based on the results of the expert's assessment, are then tested. To determine psychometric characteristics, validity, discriminatory and reliability tests were carried out. The item validity analysis technique uses the biserial correlation test because the Thurstone scale produces dichotomous data, namely the creation of two categories in the form of 1 and 0. The calculated data is said to be valid if r_{count} is greater than r_{table} (Yusuf, 2017).

Testing the differentiating power of items aims to see whether the developed instrument items are able to distinguish which groups have high religious moderation values and which individuals have low religious moderation values. Then, reliability testing uses the Kuder Ricardson 20 formula (KR20).

RESULTS

The number of items was 87 items, then a review and validation of the contents was carried out by experts, totaling 10 people. Of the 87 statement items, 27 items had a large interquartile value of 4 (> 4), so they were not significant, and 60 items had a small interquartile value of 4 (≤ 4), which meant they were significant. To find out the characteristics of the items developed, inclusion validity index V, called Aiken, was used to calculate inclusion validity based on expert judgment, and developed with indicator specifications. If the value of V is less than 0.4 (< 0.4) then the level is low, if the value of v is small than 0.8 (< 0.8) the level is moderate and if it is large than 0.8 (≥ 0.8) the level of content validity is high. With the following results:

Table 1. Validity Test Results

Item	Validity	Category	Item	Validitas	Category
1	0,920	High	47	0,700	Medium
2	0,910	High	48	0,270	Low
3	0,550	Medium	50	0,870	High
7	0,770	Medium	51	0,870	High
8	0,550	Medium	52	0,120	Low
9	0,810	High	53	0,770	Medium
12	0,870	High	54	0,630	Medium
13	0,850	High	57	0,290	Low
15	0,870	High	58	0,220	Low
16	0,610	Medium	59	0,900	High
17	0,880	High	60	0,120	Low
18	0,850	High	62	0,730	Medium
19	0,870	High	63	0,220	Low
20	0,860	High	64	0,260	Low
21	0,820	High	65	0,840	High
22	0,870	High	67	0,210	Low
23	0,860	High	70	0,650	Medium
25	0,840	High	73	0,870	High
26	0,870	High	74	0,760	Medium
27	0,860	High	75	0,220	Low
28	0,870	High	76	0,410	Medium
30	0,740	Medium	77	0,820	High
31	0,800	High	78	0,600	Medium
33	0,820	High	79	0,850	High
34	0,470	Medium	80	0,760	Medium
36	0,770	Medium	82	0,790	Medium
40	0,710	Medium	83	0,770	Medium
41	0,830	High	84	0,780	Medium
43	0,820	High	85	0,700	Medium

44 0,790 Medium 87 0,680 Medium

Based on Table 1, the category validity of the items is as follows:

Table 2. V-Aiken Validity Results

V Value	Validity Criteria	Item Number	Total
≥ 0,81	High	1, 2, 9, 12, 13, 15, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 31, 33, 41, 43, 50, 51, 59, 65, 73, 77, 79	28
0,41 – 0,80	Medium	3, 7, 8, 16, 30, 34, 36, 40, 44, 47, 53, 54, 62, 70, 74, 76, 78, 80, 82, 83, 84, 85, 87	23
< 0,40	Low	48, 52, 57, 60, 63, 64, 67, 58, 75,	9
Total			60

The results obtained are the level of content validity with a low level of 9 items, a medium level of 23 items and a high level of 28 items. The overall statement item validation level is 0.693 (<0.8 and >0.4), which means it is at a moderate level.

On the Thurstone scale the selected item should represent a scale value ranging from 1-11 (Scale attached). The arrangement of this scale point is used to assess the high moderation of students' religious attitudes for what is stated by the

statement items (Azwar, 2013). From statements that do not show an attitude of religious moderation to statement items that show an attitude of religious moderation. The continuum of religious moderation that is increasingly to the right has a high religious moderation attitude and a low moderation attitude if the value is increasingly to the left. The next step is to calculate the psychometric characteristics in terms of the validity and reliability of the scale. Validity results as follows:

Table 3. Psychometric Trial Results Validity with biserial correlation

Indicators	Description		Total items
	Valid	Not valid	
National Commitment	6	8	14
Tolerance	12	4	16
Anti violence	9	8	17
Accommodating to local culture	9	4	13

36 24 60

If $r_{count} > 0.176$ the statement is declared valid, and if $r_{count} \leq 0.176$ the statement is declared invalid. All of the statement items in Table 2, of the 60 items tested, there were 36 "Valid" items and 24 "Invalid" items. So

the statement items measure the attitude of student religious moderation into 36 items. After getting the number of valid items according to indicators, the next analysis is the analysis of differential power, as follows:

Table 4. Item Different Power Test Results

Distinction Index	Differentiation Criteria	Item	Total
$\geq 0,70$	Very good	12, 17, 27, 29, 32, 33, 36, 38, 39, 41, 43, 45, 46, 56, 57	15
0,30 – 0,39	Good	6, 7, 8, 11, 16, 20, 21, 34, 42, 48, 53, 54	12
0,20 – 0,29	Fair	10, 22, 23, 52, 55	5
$< 0,19$	poor	49, 51, 59, 60	4
	Total		36

Based on the results of the table above, a large score equal to 0.07 is in the Very Good category with 15 items, Scores in the good category are 12 items, in the sufficient category are 5 items, and in the bad category are 4 items (Azwar, 2013). Scale Statement attached.

Leon Thurstone (Yusuf, 2017) in its development has had a major impact on the development of the social sciences, especially those related to the measurement of non-cognitive aspects such as interests, attitudes and values (Setiawati et al, 2013; Zufriani et al, 2022).

The results of the reliability analysis using the KR-20 formula were assisted by the SPSS program and also tried to compare the calculation results with the Microsoft Excel program. The calculation result is 0.799 and the calculation of Microsoft Excel program is 0.800. Based on the results of this coefficient, the reliability has been satisfactory.

The Thurstone scale commonly used in attitude studies. A scale involving experts or judgments to assess each statement item (Guffey et al, 2007), responses protected from social desirability (dishonest responses) and faking (deliberately changing self-characteristics) (Chernyshenko et al, 2009; McDonald, 2013), as well as giving the opportunity to the subject to choose only one statement that is appropriate and not appropriate from the several statements provided (McDonald, 2013).

DISCUSSIONS

The instrument developed in this study was the Thurstone scale to measure students' attitudes of religious moderation. The scale originally developed by Louis

The purpose of expert judgment according to Fung (2014) is to consider the

relevance of the contents of documents that consist of several domains and the representativeness of the contents of an item to represent the domain as a whole. For example the suggestions received by Zufriani et al (2022) from experts judgment for the use of scales with agree and disagree options, because the use of choices such as (strongly agree, agree, neutral, disagree, strongly disagree) causes informants inclined to choose neutral options to avoid problems.

The construction of the Thurstone scale uses the equal-appearing interval method, which aims to sort respondents based on certain characteristics. Respondents asked to determine their attitude from very unfavorable to very supportive (favorable). Items that have been given consideration by experts are then followed by calculating frequencies, percentages, cumulative percentages, and calculating S-scores (median) and Q-scores (quartiles). The consistency value of experts can be known through the interquartile value (Q). Generically, if the Q index reaches number 4 or more, it can be said that there is an inconsistency in the evaluation given by the judge. There were 27 items that had a large interquartile value of 4 (> 4), so they were not significant, and 60 items that had a small interquartile value of 4 (≤ 4), they were declared significant. This means that of the 87 item statements developed, there are 27 statement items that are considered unfavorable by the experts (judgment). Items that are not significant are not used because the variability of the assessment is too large (Azwar, 2013). After all procedure completed, the remaining number of statement items is 60 statements with the following distribution, 14 items measuring national commitment, 16 items measuring tolerance, 17 items measuring anti-violence

and 17 items measuring accommodative towards local culture.

The selected items are significant, should be represented from numbers 1-11. The value or point of the scale is determined by the size of the tendency of the central median, so each item needs to find the median using the median formula which is given the symbol S. The results show that the distribution of the points on the scale is in the range of values 1-11. This means that the selected statement items have various scale points. The continuum of religious moderation is represented by a value that is increasingly to the right (towards the number 11) which means the higher the attitude of religious moderation and to the left (towards the number 1) means the level of moderation is low.

The development of a measuring instrument that involves expert judgment or judgment on the instrument being developed, it is recommended to use the calculation of the coefficient V or also known as Aiken validation (Aiken, 1985). V Aiken is a type of content effectiveness that is estimated by testing the feasibility or relevance of testing content through reasonable analysis by competent professionals or expert judgment (Hendryadi, 2017). Content validity is a key consideration in the measurement construction (Rudge et al, 2009). Azwar further explained the purpose of using the coefficient V to determine the extent to which the item represents the construct being measured (Azwar, 2013).

The results of the content validation test using the V coefficient for 10 expert judgments yield results with a calculated V value of 0.693. The possible range of V numbers is between 0 and 1.00 so that the number 0.693 can be interpreted as a medium coefficient which is in the range

<0.8 and >0.4 . The interpretation is high if the calculated value of V is greater than 0.8 (Aiken, 1985). V Aiken analysis to test the level of content validity has been widely used in the development of instruments such as scales.

The content validation results will generate a variety of range values, which can be factored in by the limited range of data, which can be due to test takers having similar scores so that the data variation is low, bias and measurement variation are used as criteria. Coaley (2014), decrease in the value of a statistic due to the loss of pure associations between measurement constructs (Osterlind, 2006), content, number of items and justification by the experts who gave the assessment (Mutalazimah et al., 2014).

The scale that has been assessed by the judges by default has no provisions for re-testing the respondents. However, it is necessary to conduct trials to obtain instrument items that have good psychometric characteristics since the quality of a test instrument is strongly influenced by the items, it is necessary to carry out a psychometric analysis to ensure that the test items are still functioning properly (Davidshofer & Murphy, 2005). The psychometric characteristics is related to the attributes of psychological tests. Psychometric characteristics are various characteristics related to attributes about the test (Furr & Bacharach, 2008). Various attributes of psychological tests include: the type of data or measurement result score, the reliability of the measurement, and the validity of the measurement result data.

The psychometric approach is a methodology used to address measurement problems. Kaplan and Saccuzo (2017) stated that items in a good

test is a good measurement. A number of factors, including validity, reliability, item difficulty index, item discrimination index, and distractor effectiveness, can be analyzed to determine the quality of an item.

The first step in analyzing psychometric characteristics is to conduct an empirical validity test. Completed by IAIN Metro students, with the Thurstone Scale format, namely by giving an answer "Yes" or "Agree" and an answer "No" (McDonald, 2013; Mikha, 2013; Yusuf, 2017). The number of statements tested was 60 items, and measured four aspects of religious moderation.

The validity analysis used the biserial correlation test, because the data collected is dichotomous (Mikha, 2013). Based on the analysis of the results, 36 valid and 24 invalid items were obtained. These results based on the suitability of r count and r table. If r count is greater than r table for $n = 135$ people, 0.176, then it is declared valid. Items that are invalid are removed (Yusuf, 2017). Thus, the total instrument items is 36 items, consisting of 5 items measuring aspects of national commitment, 10 items for tolerance, 11 items for non-violence and 10 items for accommodating local culture.

The results of the validity analysis have different conclusions on the psychometric validity of the content validity test. For example, in item numbers 9, 13, 15, 19, 20, 25, 31, 43, 59, 62, 65, 77, the results of content validation based on the value of V are in the high category, while the analysis using the biserial correlation test makes the item invalid or small r count of 0.176 r table. This may occur in psychometric analysis validation tests depending on the distribution of standardized sample scores (Rahayuni et

al., 2015), also could be because the person is sick, tired, in a hurry, not interested, forced, and so on (Azwar, 2013). Coaley (2014) mentioned several factors that can affect the validity of the measurement, such as similar scores from respondents so that the data is less diverse, low, relatively few respondents, decreased pure association between measurement components, and Osterlind (2006) also added that it could be due to the attenuation effect, the decrease in the value of a statistic due to the loss of pure associations between measurement constructs.

After getting the number of valid items according to the measuring aspect, then the analysis of discriminating power and reliability of the scale that has been tested. The results of the analysis of 36 statement items based on the index are 27 items in the good/satisfactory category with a calculated value above > 0.3 , there are 9 items that need to be revised. The discriminating power test aims to find out whether the developed measurement can distinguish the abilities of subjects or individuals who have a high and low attitude of religious moderation. Based on the calculation above, it can be said that the medium developed instrument has good discriminating power, considered good/satisfactory is ≥ 0.3 according to Ebel's evaluation (Azwar, 2013).

The results of the instrument reliability analysis also showed a reliability coefficient value of 0.799 or 0.800 based on the calculation of the excel program. Prayitno explains reliability less than 0.6 (<0.60) the category is not good, medium > 0.7 is acceptable and above > 0.8 is good (Azwar, 2013). The reliability index indicates that the thurstone scale instrument developed is good and can be

used as an reliable measurement. Davidshofer & Murphy (2005) stated that if the instrument items are bad, then the validity and reliability of the test would be disrupted. High reliability is necessary for tests that are used to make a final judgment for a person or to identify different categories based on their differences.

After all the statement items were analyzed based on psychometric characteristics such as validity tests, item differentiating power and also reliability tests, the results showed that the moderate religious moderation scale was acceptable and can be used to measure students' attitudes of religious moderation. When viewed from the procedures and steps for the development of the Thurstone scale, it can be said that it fulfil the requirements.

The first step in the validation is to solicit the judgment of experts to assess the content relevance. Aiken's V results show that the validation of the contents is good, there is discriminating power and reliable. Thus, indicating which item contents can express high, medium and low moderation attitudes is obtained. The items that have been selected are 36 items, arranged randomly in a scale format with the answers "yes" and "no". The process of scoring is given to the answer "yes". Responses that answered "Yes" will be given a value equal to the value of S. Respondents who answered "No", then not given a score or value of 0. Then carry out the interpretation by calculating the "total score", then calculating the mean of the S value, then the mean value (average) put it on a scale range 1-11 which is the subject position for the variable measured.

The result that is further to the right (towards number 11) means that the attitude of religious moderation is getting higher and to the left (towards number 1)

means that the level of moderation is getting lower (attached). The magnitude of the value of an item statement or a response based on the answer choices. Positive answer to favorable item is treated the same as negative answer to unfavorable item, which given a high score. Conversely, negative answer to favorable item is treated in the same way as positive answer to unfavorable item, which is given a low score (Azwar, 2013).

Additionally, the Thurstone method have advantages, thus suspected better in measuring students' religious moderation attitudes. This scale provides scores according to a clearer ranking because each aspect is seen in a larger context. It also leverages the expert opinion obtained by calculating the differences between each aspect of the interval scale, providing more complete information than using a simple scale. The yes and no answer option in measuring the attitude of religious moderation are able to avoid untruthful answer. In addition, using the rule, each stimulus is paired with another stimulus, and a certain number of items is obtained such that all items are paired with each other. Expectantly, the religious moderation in Thurstone scale will be able to provide a consistent response in each item. Religious moderation scales with options Yes and No or Discrete items are able to explain individual performance more reliably than Likert-scale items that likely uncertain and respondent may give a slightly misleading answer. Not all discrete statements stand up to tricks in answers. The dichotomous item succeeds in limiting misleading responses if the two loaded alternatives are equivalent in terms of quality weight.

The development of the Thurstone scale to measure student religious

moderation attitudes have several limitations. The development was carried out in 2021 during the Covid-19 pandemic that the involvement of experts was carried out online and on a Google form which was not directly confirmed (face to face). Sometimes there is a misunderstanding of intent in communicating. Then the indicators and definitions used in this study are also accepted by the majority of experts, although there are still those who view the concept of religious moderation as not having an operationalization of religious moderation behavior with a clear cut. Furthermore, it is important that the items on each indicator mapped in advance based on the cognitive-affective-psychomotor aspect or it can also be the perspective-attitude-behavior which is an important suggestion for further research.

CONCLUSIONS

The religious moderation scale to measure students' religious moderation attitudes fulfilled the psychometric element means that the scale of religious moderation can be used as an instrument to measure students' religious moderation. The total of 36 items represents the 4 indicators of religious moderation; national commitment, tolerance, anti-violence and accommodation of local culture. Based on the analysis results of the psychometric characteristics based on item indices' validity, reliability, and discriminating power, the authors confirmed that the instrument could identify the level of religious moderation with consistent results and can be used at other times in related studies.

REFERENCES

Abdi, A. P. (t.t.). *Setara Institute Sebut 10 Kampus Terpapar Paham Radikalisme*. tirtoid. Diambil 7 Mei

- 2021, dari <https://tirto.id/setara-institute-sebut-10-kampus-terpapar-paham-radikalisme-d9nh>
- Abdullah, M., Nasution, A., Siregar, Abd. A., & et all. (2020). Literasi Moderasi Beragama di Indonesia. Bengkulu: Zigie Utama.
- Abu-Raiya, H., & Hill, P. C. (2014). Appraising the state of measurement of Islamic religiousness. *Psychology of Religion and Spirituality*, 6(1), 22–32. <https://doi.org/10.1037/a0035082>
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45, 131–142. <https://doi.org/10.1177/07399863870092005>
- Ali, H., & Purwandi, L. (2016). Indonesia 2020: The Urban Middle Class Millenials. Jakarta: PT Alvara Strategi Indonesia.
- Amir, Y. (2021). Pengembangan Skala Religiusitas untuk Subyek Muslim. *Indonesian Journal for The Psychology of Religion*, 1(1), Art. 1. <https://doi.org/10.24854/ijpr403>
- Antara. (2020, Januari 18). Survei Wahid Institute: Intoleransi-Radikalisme Cenderung Naik. <https://mediaindonesia.com/politik-dan-hukum/284269/survei-wahid-institute-intoleransi-radikalisme-cenderung-naik>
- Aslam, M. M., & Gunaratna, R. (2019). Terrorist rehabilitation and community engagement in Malaysia and Southeast Asia. Routledge London.
- Azra, A. (2003). Bali and Southeast Asian Islam: Debunking the myths. After Bali: The Threat of Terrorism in Southeast Asia, 39–57.
- Azra, A. (2017). Islam Indonesia Inklusif vs Eksklusif: Dinamika Keberagaman Umat Muslimin. Makalah untuk Pengajian Ramadhan PP Muhammadiyah Kampus Universitas Muhammadiyah Jakarta, 6.
- Azwar, S. (2013). Penyusunan Skala Psikologi. Jakarta: Kencana.
- Barker, E. (2022). New Religious Movements. Dalam L. R. Kurtz (Ed.), *Encyclopedia of Violence, Peace, & Conflict (Third Edition)* (hlm. 339–353). Academic Press. <https://doi.org/10.1016/B978-0-12-820195-4.00286-7>
- Chernyshenko, O. S., Stark, S., Prewett, M. S., Gray, A. A., Stilson, F. R., & Tuttle, M. D. (2009). Normative Scoring of Multidimensional Pairwise Preference Personality Scales Using IRT: Empirical Comparisons with Other Formats. *Human Performance*, 22(2), 105–127. <https://doi.org/10.1080/08959280902743303>
- Coaley, K. (2014). An Introduction to Psychological Assessment & Psychometrics (2nd edition). SAGE.
- Davidshofer, K. R., & Murphy, C. O. (2005). *Psychological testing: Principles and applications*. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Furr, R. M., & Bacharach, V. (2008). *Psychometrics and the importance of psychological measurement*. Psychometrics. Thousand Oaks, CA: Sage Publications Inc.
- Guffey, J. E., Larson, J. G., Zimmerman, L., & Shook, B. (2007). The development of a Thurstone scale for identifying desirable police officer traits. *Journal of Police and Criminal Psychology*, 22(1), 1–9.
- Hadiz, V. R. (2016). *Islamic populism in Indonesia and the Middle East*. Cambridge University Press.
- Hasanuddin, A. (t.t.). » Ideologi Negara: Pancasila atau Islam? - Alvara

- Strategic Ideologi Negara: Pancasila atau Islam? Diambil 6 Mei 2021, dari <https://alvara-strategic.com/ideologi-negara-pancasila-atau-islam/>, <https://alvara-strategic.com/ideologi-negara-pancasila-atau-islam/>
- Hendryadi, H. (2017). Validitas Isi: Tahap Awal Pengembangan Kuesioner. *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT*, 2(2), 169–178. <https://doi.org/10.36226/jrmb.v2i2.47>
- Jabali, F., & Jamhari, I. (2002). *Modernisasi Islam di Indonesia* (Jakarta). Logos. *Anales Del Seminario de Metafísica [Universidad Complutense de Madrid, España]*.
- Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological testing: Principles, applications, and issues*. Cengage Learning.
- Kholis, N., Azra, A., Hasan, N., Qodir, Z., Qibtyah, A., Sadzali, A., & Min Fadhl Robby, H. (2020). *Islam Indonesia 2020*. UII Press Yogyakarta.
- Krabbe, P. F. M. (2008). Thurstone Scaling as a Measurement Method to Quantify Subjective Health Outcomes. *Medical Care*, 46(4), 357–365. JSTOR.
- Lisnawati, S. (2011). Development Of Emotional Quotient Instrument. *Jurnal Evaluasi Pendidikan*, 2(1), 54. <https://doi.org/10.21009/JEP.021.05>
- McDonald, R. P. (2013). *Test theory: A unified treatment*. psychology press.
- Meiza, A. (2018). Sikap Toleransi dan Tipe Kepribadian Big Five pada Mahasiswa UIN Sunan Gunung Djati Bandung. *Psychopathic : Jurnal Ilmiah Psikologi*, 5(1), Art. 1. <https://doi.org/10.15575/psy.v5i1.1959>
- Mikha, A. W. (2013). *Statistika Terapan; Konsep dan Aplikasi SPSS/LISREL dalam Penelitian Pendidikan, Psikologi dan Ilmu Sosial Lainnya*. Jakarta: PT Elex Media Komputindo.
- Munir, M. B., & Herianto, H. (2020). Tingkat Pemahaman Moderasi Beragama Serta Korelasinya Terhadap Pengaruh Kesehatan Mental, Keaktifan Berorganisasi dan Prestasi Akademik. *Prosiding Nasional*, 3, 137–150.
- Murtadlo, M. (2020). Menakar Moderasi Beragama di Perguruan High. Badan Litbang dan Diklat Kemenag RI. <https://doi.org/10.31219/osf.io/jg8hx>
- Mutalazimah, -, Azwar, S., & Murti, B. (2014). Penskalaan Thurstone pada Aitem Thyroid Dysfunction Questionnaire (TDQ) Berbasis Gejala Biopsikososial. *Jurnal Psikologi*, 41(2), Art. 2. <https://doi.org/10.22146/jpsi.6945>
- Nasir, M., & Rijal, M. K. (2021). Keeping the middle path: Mainstreaming religious moderation through Islamic higher education institutions in Indonesia. *Indonesian Journal of Islam and Muslim Societies*, 11(2), Art. 2. <https://doi.org/10.18326/ijims.v11i2.213-241>
- Osterlind, S. J. (2006). *Modern measurement: Theory, principles, and applications of mental appraisal*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- PPIM UIN Jakarta. (2021). *Potret Moderasi Beragama di Kalangan Mahasiswa Muslim :Kasus Tiga Kampus Islam* (Jakarta, Bandung, Yogyakarta). Pusat Pengkajian Islam dan Masyarakat (PPIM) UIN Jakarta.
- Pratama, D. (2020). Pengembangan Skala Thurstone Metode Equal Appearing Interval untuk Mengukur Sikap Moderasi Beragama Siswa Sekolah Menengah Atas. *Jurnal Psikologi Teori dan Terapan*, 11(1), 71.

<https://doi.org/10.26740/jppt.v11n1.p7>
1-82

- Rahayuni, K., Yunus, M., & Fadil, R. (2015). Analisis Psikometri Validitas Dan Reliabilitas Kuesioner Athletic Coping Stress Inventory-28 (CSAI-28) Dalam Bahasa Indonesia. *Motion: Jurnal Riset Physical Education*, 6(1), 17–31.
- Rijal, M. K., Nasir, M., & Rahman, F. (2022). Potret Moderasi Beragama di Kalangan Mahasiswa. *PUSAKA*, 10(1), Art. 1. <https://doi.org/10.31969/pusaka.v10i1.672>
- Rudge, C. V., Calderon, I. M., Dias, A., Lopes, G. P., Barbosa, A. P., Maestá, I., Odland, J. Ø., & Rudge, M. V. (2009). Design and validity of a questionnaire to assess sexuality in pregnant women. *Reproductive Health*, 6, 12. <https://doi.org/10.1186/1742-4755-6-12>
- Schmid, A. P. (2015). Challenging the narrative of the “Islamic State.” *The Hague: International Centre for Counter-Terrorism*, 14.
- Setiawati, F. A., Mardapi, D., & Azwar, S. (2013). Penskalaan Teori Klasik Instrumen Multiple Intelligences Tipe Thurstone dan Likert. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 17(2), Art. 2. <https://doi.org/10.21831/pep.v17i2.1699>
- Yusuf, A. M. (2017). *Metode Penelitian Kuantitatif, Kualitatif dan Penelitian Gabungan* (4 ed.). Kencana.
- Zufriani, Z., Pitriani, P., & Damni, A. (2022). Rasch analysis of student attributes: Development and validation of scale to measure religious moderation. *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, 8(1), 187–195.

