Rasch analysis of student attributes: development and validation of scale to measure religious moderation

Zufriani Zufriani[*)], Pitriani Pitriani, Albertos Damni
Institut Agama Islam Negeri Kerinci, Indonesia

Article Info

ABSTRACT

In Indonesia, the narrative of Religious Moderation is being hotly discussed and has even been included in the RPJMN 2020-2024. The target index has been determined, but measurement instruments are still minimal and can be used to measure the level of religious moderation at the university level. This research was conducted in response to these needs. The development of this instrument uses a unidimensional construct using Rasch analysis. The number of samples used was 406 students selected using the multistage random sampling technique. This development goes through three stages, namely item generation,where the objectives and concepts of measuring religious moderation are arranged to produce latent items. The theoretical analysisaims to test the scale's content validity.The last stage is a psychometric analysiswhich seeks to analyze the availability of the sample and construct the validity and reliability of the scale using Rasch analysis. The first stage produced 48 statements; thesecond stage resulted in changes in the religious context and usedmore appropriate words to describe the scale interval. The last stage makes 24 valid argumentsto be sureof religious moderation.

Corresponding Author:

Zufriani Zufriani,
Institut Agama Islam Negeri Kerinci
Email: zufrianistainkerinci@gmail.com

Introduction

One of Indonesia's increasingly discussed narratives is Religious Moderation/Wasathiyah (Arif, 2020; Bagi et al., 2021; kementerian RI, 2019; Sutrisno, 2019). This narrative is considered one of the potential solutions to maintaining integrity and harmony, both inter-religious and intra-religious. This narrative is also used as a counter-narrative against the rampant ideology of Pancasila rejection, Radicalism, and acts of intolerance in Indonesia (Islam & Bumi, 2016; Supriadi et al., 2020; Untoro & Putri, 2019).

This narrative has been included in one of the National Character Developments in the 2020-2024 National Medium-Term Development Plan (RPJMN). The four main character-building points are Religious Moderation; Mental Revolution and Pancasila Ideology Development; Promotion and Preservation of Culture; and Culture of literacy, Innovation, and Creativity.

Each achievement target index is more measurable, and there is no Dunning-Kruger effect; those with a low achievement index feel high and vice versa(Schlösser et al., 2013). For example, the mental revolution achievement index is targeted at 74.2, cultural literacy at 71.0, and religious harmony at 75.8. However, the achievement index of the Religious Moderation target is still biased because it is not given an achievement target that should be achieved in 2024.

This achievement target in 2024 was developed through one of the programs, namely the Mainstreaming of Religious Moderation. At universities under the Ministry of Religion, PTKIN must establish a house of religious moderation based on a letter circulated Number B-3663.1/Dj.I/BA.02/10/2019 concerning the Circular of Religious Moderation Houses issued by the Directorate General of Education, Ministry of Religion.

In determining the target index for achieving religious moderation, a national standard is neededto reflect Indonesia's level of religious moderation.Religious moderation should be assessed as a way of looking at the country from a religispiritual of view. Religious moderation should also be able to give birth to an anti-extreme attitude toward a religious ideology. Furthermore, religious moderation must go beyond the formalization of religion, such as ritual and on-ritual worship(Barker, 2022; Berkes, 2017; Juergensmeyer, 2015).

The target of achieving religious moderation should be indexed and stated in the form of valid and reliable measuring instruments (instruments). This is intended so that it can be used on a broad scale and be able to become an appropriate benchmark for achieving the 2020-2024 RPJMN targets.

Instruments for measuring religious moderation have indeed been made. Still, they have not been able to fully measure the level of good religious moderation, and some only measure aspects of religiosity and community tolerance(Abu-Raiya & Hill, 2014; Amir, 2021; Ersanli & Mameghani, 2016). However, the instrument still has limitations, such as without using Target Population Judges (TPJ)(Cormack, 2001).

The researchers tried to develop these limitations in this follow-up study by following a more valid stage by the rules of a suitable instrument development method(Clark & Watson, 1995; DeVellis, 2003; Meneses et al., 2013). This study aims to develop an instrument of religious moderation using three stages of development, namely item generation; the researcher uses the deduction method, which uses literature review to produce items. Theoretical Analysis stage, the researcher combines expert opinion and opinion from the target population of the research sample. And the last stage is Psychometric Analysis; the researcher uses Rasch Analysis to validate the accuracy of the items(Bartholomew, 2015; Junker, 2015; Van Der Linden, 2010).

Method

Researchers in this study tried to develop a measurement instrument. Therefore, the researcher uses a research and development approach using three stages, item generation, theoretical analysis, and psychometric analysis. The population in this study were all IAIN Kerinci students registered in the 2020/2021 Academic Year. At the same time, the sample is taken by considering the number of members in the components in the population. The sample selection technique will use the Multistage Cluster Random Sampling technique, a sampling technique that makes the elements in the population as units with specific characteristics from individuals that then function as population representatives (Cresswell, 2009; Nursiyono, 2015). These population characteristics refer to the demographic variables of the population. Because the population in this study is all students of IAIN Kerinci, the attributes in this sample selection technique also refer to demographic variables, namely faculty, department, and semester level. The total number of initial samples to be taken amounted to 360 people. The consideration is that the number of initial items in the development of this instrument is about 19 items, and the ratio of items and respondents used is 1: 15. The number of samples in this study still needs to be added with a backup selection to anticipate the scale distribution is not filled in correctly, is not returned, or is scattered. The number of reserve samples in this study was around 20% of the model, considering cost and workforce. So that the total sample after being added to the reserve sample in this study was 432 students. In the early stages, the deduction method was used through a literature review which then combined the results from expert judges with the target population (Target Population Judges,) and finally, the data analysis used psychometric analysis and was analyzed using Rasch analytical modeling, which is one of the analytical models thatemphasisess the principle of objective comparison (Andrich et al., 2011). In the Rasch analysis, the modeling is analyzed during the psychometric analysis stage, namely the validity and reliability of the instrument. Instrument analysis is expected to eliminate, improve and maintain good and reliable items on the religious moderation measurement scale.

Results and Discussions

Item Generation
It is making items in the research on developing a religious moderation scale using a literature review as the first step. The Religious Moderation Scale was created using the unidimensional framework,namely the creation of items that use one dimension without being manifested by sub-constructs or construct dimensions (Andrich, 2010; Sumintono & Widhiarso, 2014).

Theoretical Analysis
This research uses two steps at the theoretical analysis stage: asking for expert opinions (Expert Judges) and ideas from the target population (Target Population Judges).

Expert Judges
Instrument experts suggest replacing the use of a "neutral" scale which is worth 3 in the interval range (strongly agree, agree, neutral, disagree, strongly disagree) to be replaced with the use of the word "between agree and disagree". The use of neutral words can cause informants to be more inclined to choose that option. This is because humans prefer neutrality to avoid problems(Dubow & Rubinlicht, 2011; Matthews et al., 2015). The second expert suggested that the religious context in the scale should be reproduced because this scale is to be used in religious moderation in the future.

Target Population Judges
In some cases, students did not understand the context of statements such as the use of the word demonstration, lectures, and television shows. For demonstrations, students don't understand the flow that must be done to carry out demonstrations; forclasses, students tend to discredit the use of the word only for Muslims in the context of da'wah, and now students rarely watch television because they feel smartphones have replaced them. Because of these problems, the researcher tried to improve the words to make them easier to understand as the context of the statement in giving consentbecause the context of language is very important in perceiving something(Minakova & Gural, 2015; Willems & Peelen, 2021).

Psychometric Analysis
Trial Phase
The measurement trial data for the religious moderation scale are 26 valid samples from 26 total samples and 48 valid initial items from 48 total. Standardized Residual N(0,1), Mean 0.01 and Standard Deviation 1.00. Data is normally distributed as indicated by an average value of 0.01 (which means close to 0) with a standard deviation of 1.00 (the closer to 1, the more valid))(Andrich, 2010; Bond & Fox, 2013).

Table 1. Validity and Reliability Trial Phase

Label	Category	Infit		Outfit		Reliability	Separation
		MNSQ	ZSTD	MNSQ	ZSTD		
Person	Mean	1.01	-0.30	1.00	-0.20	0.75	1.72
	S.D	0.50	2.40	0.47	2.00		
	Max	2.67	6.10	2.65	5.40		
	Min	0.28	-5.30	0.39	-3.80		
Item	Mean	1.01	0.00	1.00	0.00	0.93	3.78
	S.D	0.05	0.31	1.10	0.33		
	Max	1.91	3.00	1.97	3.20		
	Min	0.48	-2.30	0.48	-2.40		

Tabel 1 shows the validity and reliability of the measurement. Person Reliability in this measurement is relatively high, with a value of 0.75. This can indicate a measuring tool that is sensitive enough to separate those who have high and low abilities (Andrich, 2005; Bond & Fox, 2015).However, the minimal Separation value is 1.72, the mean infit and outfit values are excellent (MNSQ, which is close to 1, almost comparable to 0), and the mean measure value is 0.0; this indicates a relatively similar range of sample abilities. When measured using CTT (Classical Test Theory), this reliability value is also classified as quite good, marked by a measuring value of 0.78.

Item Reliability in this measurement can be excellent, with a measuring value of 0.9; the Infit value and MNSQ and ZSTD outfits are also perfect. The average MNSQ infit is 1.01, the average outfit is MNSQ 1.00, the ZSTD infit value is 0, and the ZSTD outfit is 0.

All items can be categorizedinthe perfectordered table above. The most positive PMC values indicate this (in one measurement direction), and the Outfit and Infit MNSQ values are all in the range of 0.5-1.5. In the next stage, PMCs with negative values cannot be included in the wider instrument distribution stage because they fear they will not be in the same measurement direction (27, 30, 13, 38, 42, 33, and 11). The distribution of the choice of answers from respondents can be seen in full in Table 1

Table 2. Fit Statistics Trial Phase

Entry	Measure	In.msq	Out.msq	Pmc
1	1.20	0.94	0.81	0.51
2	-1.19	1.21	1.14	0.07
3	-0.20	1.06	1.01	0.61
4	-0.32	1.49	1.41	0.60
5	0.65	0.80	0.78	0.45
6	-0.36	1.05	1.04	0.48
7	0.28	1.40	1.37	0.48
8	-0.93	0.77	0.64	0.57
9	0.00	0.99	1.03	0.30
10	-0.57	0.73	0.66	0.52
11	-0.61	0.67	0.61	-0.03
12	-0.66	1.26	1.20	0.22
13	-1.42	0.78	0.73	-0.16
14	-0.87	0.63	0.62	0.20
15	0.69	1.05	1.06	0.16
16	1.29	0.48	0.48	0.45
17	0.78	0.85	0.91	0.20
18	0.62	0.80	0.82	0.49
19	0.62	0.73	0.73	0.35
20	0.78	0.73	0.72	0.21
21	-0.48	0.89	0.83	0.60
22	0.85	1.06	1.11	0.16
23	0.82	1.09	1.21	0.07
24	1.29	0.76	0.74	0.09
25	-0.13	0.80	0.76	0.58
26	-0.10	0.85	0.82	0.30
27	0.99	1.30	1.52	-0.37
28	-1.68	1.91	1.97	0.16
29	0.28	0.96	0.96	0.44
30	-0.66	1.04	1.01	-0.20
31	-1.50	1.44	1.43	0.01
32	-0.61	1.43	1.24	0.35
33	-1.89	1.25	1.25	-0.07
34	1.62	0.66	0.53	0.51
35	-0.13	0.51	0.48	0.63
36	-0.40	1.08	1.24	0.29
37	0.59	1.76	1.83	0.19
38	0.44	1.02	1.06	-0.09
39	-1.68	0.66	0.64	0.11
40	0.03	1.11	1.10	0.54
41	1.92	0.61	0.66	0.03
42	0.69	0.96	0.95	-0.08
43	-0.76	1.49	1.51	0.14
44	1.38	1.06	1.02	0.21
45	-0.66	0.64	0.67	0.38
46	0.32	1.09	1.12	0.09
47	-0.24	1.21	1.18	0.63
48	-0.07	1.32	1.29	0.70

Broader Instrument Dissemination Stage

The second measurement for the religious moderation scale is406 valid samples from 407 total samples and 48 valid initial items from 48 total. Standardized Residual N(0,1), Mean 0.01 and Standard Deviation 1.00. Data is normally distributed as indicated by an average value of 0.01 (which means close to 0) with a standard deviation of 1.00 (the closer to 1, the more valid))(Andrich, 2010; Bond & Fox, 2013).

Table 3. Validity and Reliability instrument

| Label | Category | Infit | | Outfit | | Reliability | Separation |
		MNSQ	ZSTD	MNSQ	ZSTD		
Person	Mean	1.01	-0.20	1.01	-0.20	0.81	2.06
	S.D	0.50	2.30	0.49	2.00		
	Max	2.47	5.10	2.60	4.60		
	Min	0.21	-5.70	0.22	-0.50		
Item	Mean	1.01	-0.30	1.01	-0.30	1.00	15.21
	S.D	0.32	4.20	0.36	4.40		
	Max	1.88	9.90	2.05	9.90		
	Min	0.48	-9.20	0.48	-9.80		

Person Reliability in the second measurement is higher with a value of 0.81than before 0.75. This can indicate a measuring tool that is very sensitive to separate those who have high and low abilities(Andrich, 2005; Bond & Fox, 2015). The Separation value is 2.06, whichindicates there is 3 level ability person.Mean infit and outfit are perfect (MNSQ, which is close to 1, ZSTD is almost comparable to 0), and the mean measure value of -0.01 indicates a relatively similar range of sample abilities. When measured using CTT (Classical Test Theory), this reliability value is also classified as quite good, marked by a measuring value of 0.83.

Item Reliability in this measurement can be said to be very good, with a measuring value of 1.00; the Infit dealof MNSQ and ZSTD outfits are also excellent. The average MNSQ infit is 1.01, the average outfit is 1.01 MNSQ, the ZSTD infit value is -0.2, and the ZSTD outfit is -0.2.

Table 4. Fit Statistics Instrument

Entry	Measure	In.msq	Out.msq	Pmc
1	1.16	0.89	0.76	0.52
2	-1.26	1.23	1.17	0.07
3	-0.27	0.98	0.91	0.65
4	-0.39	1.36	1.26	0.66
5	0.59	0.78	0.76	0.47
6	-0.42	1.09	1.06	0.45
7	0.23	1.39	1.35	0.49
8	-1.00	0.71	0.61	0.58
9	-0.07	1.01	1.07	0.31
10	-0.65	0.69	0.60	0.55
11	-0.74	1.25	1.18	0.24
12	-0.94	0.68	0.66	0.19
13	0.65	1.11	1.17	0.15
14	1.24	0.48	0.51	0.44
15	0.74	0.92	1.06	0.17
16	0.57	0.80	0.81	0.50
17	0.57	0.76	0.76	0.34
18	0.73	0.77	0.74	0.24
19	-0.55	0.83	0.79	0.61
20	0.80	1.11	1.24	0.15
21	0.76	1.17	1.37	0.06
22	1.25	0.81	0.77	0.09
23	-0.20	0.81	0.78	0.52
24	-0.17	0.85	0.82	0.34
25	-1.75	1.87	2.05	0.14
26	0.22	0.97	0.98	0.41
27	-1.57	1.41	1.41	0.06
28	-0.68	1.44	1.22	0.33
29	1.58	0.62	0.50	0.51

Zufriani, Z. et al.   JPPI (Jurnal Penelitian Pendidikan Indonesia)
Vol. 8, No. 1, 2022, pp. 187-195

192

Entry	Measure	In.msq	Out.msq	Pmc
30	-0.20	0.52	0.48	0.61
31	-0.48	1.04	1.28	0.31
32	0.54	1.88	2.04	0.15
33	-1.75	0.68	0.66	0.13
34	-0.03	1.09	1.08	0.53
35	1.88	0.64	0.73	0.02
36	-0.84	1.44	1.46	0.18
37	1.33	1.11	1.09	0.19
38	-0.74	0.66	0.70	0.38
39	0.26	1.14	1.21	0.10
40	-0.31	1.09	1.04	0.69
41	-0.12	1.23	1.19	0.72

All items can be categorizedinthe perfectordered table above. This is indicated by the PMC values , which are all positive (in one direction of measurement), and the Outfit and Infit MNSQ values which are all in the range of 0.5-1.5(Andrich, 2005; Bond & Fox, 2013). The distribution of answer choices can be seen in the figure below
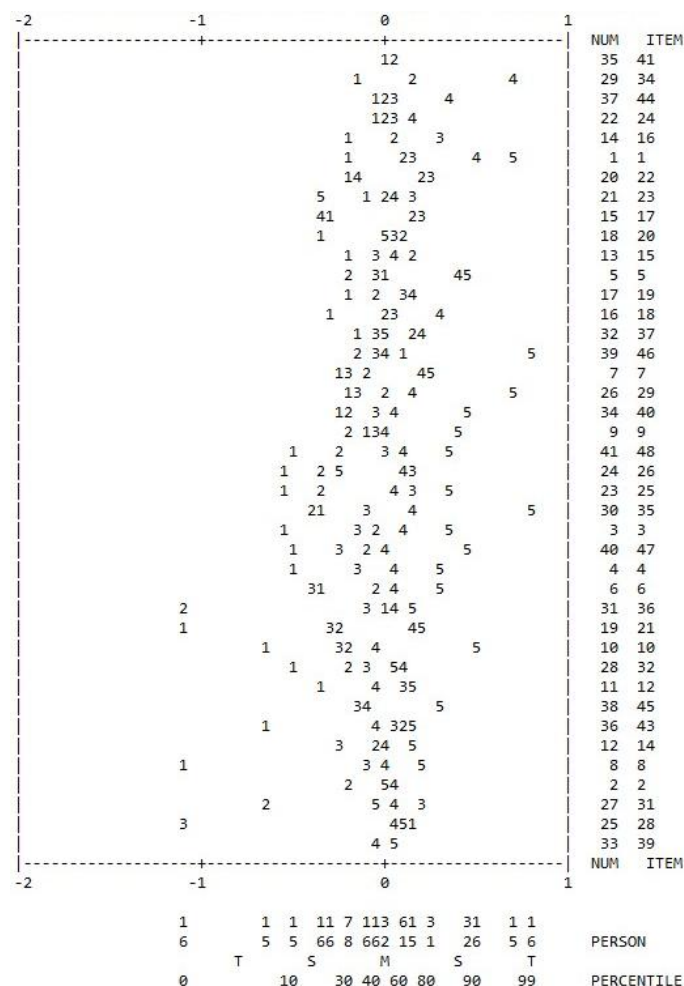

Figure 1. Observed Average Measure for Person

Overall the item categories in the measurement scale function well because of the selection of items in one measurement direction (Borsboom & Molenaar, 2015; Heise, 2015). The most accessible item to agree with is item 33, and the most complicated item to agreewith is item 35. The structure of the selection of agreement-disagreement of each respondent can be seen in the table below

Table 5. Summary of Category Structure

Category Label	Score	Observed Count	%	Observed Average	Sample Expect	Infit MNSQ	Outfit MNSQ	Category Measure
1	1	3130	19	-0.92	-1.03	1.24	1.29	-2.18
2	2	2870	17	-0.71	-0.55	0.80	0.82	-0.93
3	3	2763	17	-0.11	-0.06	0.83	0.74	-0.15
4	4	5321	32	0.50	0.45	0.91	0.94	0.81
5	5	2562	15	0.95	0.95	1.02	1.04	2.64

From the table above, it can be seen that the observed average and sample expect to have almost the same value, which indicates that the category selection of the respondents does not have a problem. To see how strong the item forms the measurement scale, then in the analysis of the Standardized Residual Variance in Eigenvalue, the complete results can be seen in the table below.

Table 6. Standardized Residual Variance (in Eigenvalue units)

```
                                          -- Empirical --    Modeled
Total raw variance in observations    =   80.8 100.0%        100.0%
  Raw variance explained by measures  =   39.8  49.3%         49.7%
    Raw variance explained by persons =    6.6   8.2%          8.2%
    Raw Variance explained by items   =   33.2  41.1%         41.4%
Raw unexplained variance (total)      =   41.0  50.7% 100.0%  50.3%
  Unexplned variance in 1st contrast  =    7.3   9.0%  17.8%
  Unexplned variance in 2nd contrast  =    5.2   6.4%  12.6%
  Unexplned variance in 3rd contrast  =    3.8   4.7%   9.2%
  Unexplned variance in 4th contrast  =    3.0   3.7%   7.3%
  Unexplned variance in 5th contrast  =    3.0   3.7%   7.2%
```

The problem Unexplained variance in 1st contrast is worth 7.3, which should not exceed 2 (Bond & Fox, 2013; Engelhard & Wind, 2019),but the variance explained by the 1st contrast is 9% smaller than item difficulties which are worth 41.1% which means that there is no second dimension in the data because Rasch item difficulties have fully explained everything.

So the final item is based on psychometric analysis after eliminatithe ng same measurement values so that items do not measure the same thing. It produces19 most valuable items. More details can be seen in the table below.

Table 7. Final Item

Item	Measure
I don't know why we have to go to the mosque for Friday prayers	-1.75
During the fasting month, my parents often force me to have sahur	-1.57
Sometimes, I get annoyed with people who come to houses to ask for donations	-1.26
I am pretty sure that the Pancasila ideology can strengthen the unity of the Indonesian state	-1.00
Pancasila is not the best ideology for the Indonesian people	-0.94
The application of Islamic Sharia will make Indonesia a better place	-0.74
I wouldn't like that we have to holidays on other people's religious days	-0.65
On December 25th, Christmas attributes should not be displayed in public	-0.42
Non-Muslims should not mix with Muslims	-0.27
People of other religions should learn about Islam.	-0.12
If my house is next to the church, I don't like the congregation parking in front of my house	-0.07
I will offer to take other people to church if I know their vehicle is breaking down	0.22
I don't expect thanks to the people I help	0.57
Calm is the same as slow	0.65
I hope our country does not cooperate with communist countries	0.80
We only need to maintain good relations with those closest to us	1.16
Students must serve the community	1.25
I don't know why we have to like cooperation	1.33
In discussions, not everyone should be given the same right to have an opinion	1.88

Conclusions

The construction of this scale goes through three stages; the first stage produces 48 unidimensional items.Thenexperts suggest replacing the use of a "neutral" scale which is worth 3 in the interval range (strongly agree, agree, neutral, disagree, strongly disagree) to be replaced with the use of the word "between agreeing and disagree" and adjusting the scale in the religious context. The researcher also tried to improve the terms to make them more accessible for the target population to understand as the context of the statement in giving consent than the scale tested on 26 samples, resulting in 41 items. The final step, hierarchy, is spread to 407 students and generatesthe 19 most valuable items that can be used to measure religious moderation.

References

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

Amir, Y. (2021). Pengembangan Skala Religiusitas untuk Subyek Muslim. *Indonesian Journal for The Psychology of Religion*, *1*(1), 47–60. https://doi.org/10.24854/ijpr403

Andrich, D. (2005). Rasch, Georg. *Encyclopedia of Social Measurement*, 299–306. https://doi.org/10.1016/B0-12-369398-5/00477-1

Andrich, D. (2010). Rasch Models. *International Encyclopedia of Education*, 111–122. https://doi.org/10.1016/B978-0-08-044894-7.00258-X

Arif, M. K. (2020). Moderasi Islam (Wasathiyah Islam) Perspektif Al-Qur'an, As-Sunnah Serta Pandangan Para Ulama Dan Fuqaha. *Al-Risalah*, *11*(1), 22–43. https://doi.org/10.34005/alrisalah.v11i1.592

Bagi, B., Milenial, G., & An, Q. U. R. (2021). *Konsep Internalisasi Nilai-Nilai Moderasi the Concept of Internalizing the Values of Religious*. *4*(1), 72–82.

Barker, E. (2022). New Religious Movements. *Encyclopedia of Violence, Peace, & Conflict*, 339–353. https://doi.org/10.1016/B978-0-12-820195-4.00286-7

Bartholomew, D. J. (2015). Factor Analysis and Latent Structure Analysis: Overview. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 691–697. https://doi.org/10.1016/B978-0-08-097086-8.42043-X

Berkes, F. (2017). Religious Traditions and Biodiversity. *Reference Module in Life Sciences*. https://doi.org/10.1016/B978-0-12-809633-8.02335-9

Bond, T. G., & Fox, C. M. (2013). Applying the Rasch Model. In *Applying the Rasch Model*. Routledge. https://doi.org/10.4324/9781410614575

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences*.

Borsboom, D., & Molenaar, D. (2015). Psychometrics. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 418–422. https://doi.org/10.1016/B978-0-08-097086-8.43079-5

Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309.

Cormack, R. (2001). Population Size Estimation and Capture–Recapture Methods. *International Encyclopedia of the Social & Behavioral Sciences*, 11809–11813. https://doi.org/10.1016/B0-08-043076-7/00486-1

Cresswell, J. (2009). *Research Design: Pendekatan kualitatif, kuantitatif dan Mixed*. Pustaka Pelajar.

DeVellis, R. (2003). *Scale Development: theory and application*.

Dubow, E. F., & Rubinlicht, M. (2011). Coping. *Encyclopedia of Adolescence*, *3*, 109–118. https://doi.org/10.1016/B978-0-12-373951-3.00107-1

Engelhard, G., & Wind, S. A. (2019). Invariant Measurement with Raters and Rating Scales. In *Invariant Measurement with Raters and Rating Scales*. https://doi.org/10.4324/9781315766829-14

Ersanli, E., & Mameghani, S. S. (2016). Construct Validity and Reliability of the Tolerance Scale among Iranian College Students. *Journal of Education and Practice*, *7*(34), 99–105. http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=eric&AN=EJ1126670&amp%0Alang=es&site=ehost-live&scope=site

Heise, D. R. (2015). Scaling and Classification in Social Measurement. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 4–8. https://doi.org/10.1016/B978-0-08-097086-8.44052-3

Islam, R., & Bumi, D. (2016). RADIKALISME ISLAM DI BUMI NUSANTARA Sukarno L. Hasyim 1. *Lentera*, *1*, 11–25.

Juergensmeyer, M. (2015). Religious Nationalism and the Secular State. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 401–405. https://doi.org/10.1016/B978-0-08-097086-8.64006-0

Junker, B. W. (2015). Factor Analysis and Latent Structure: IRT and Rasch Models. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 698–702. https://doi.org/10.1016/B978-0-08-097086-

8.42126-4

kementerian RI. (2019). *Moderasi Beragama*.

Matthews, G., Zeidner, M., & Roberts, R. D. (2015). Coping, Personality and. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 895–901. https://doi.org/10.1016/B978-0-08-097086-8.25048-4

Meneses, J., Barrios, M., Bonillo, A., Cosculluela, A., Lozano, L., Turbany, J., & Valero, S. (2013). *Psicometría*.

Minakova, L. Y., & Gural, S. K. (2015). The Situational Context Effect in Non-Language-Majoring EFL Students' Meaning Comprehension. *Procedia - Social and Behavioral Sciences*, *200*, 62–68. https://doi.org/10.1016/J.SBSPRO.2015.08.014

Nursiyono, J. A. (2015). *Pengambilan Sampel*. In Media.

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, *39*, 85–100. https://doi.org/10.1016/j.joep.2013.07.004

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial (edisi revisi)*.

Supriadi, E., Ajib, G., & Sugiarso, S. (2020). Intoleransi dan Radikalisme Agama: Konstruk LSM tentang Program Deradikalisasi. *JSW (Jurnal Sosiologi Walisongo)*, *4*(1), 53–72. https://doi.org/10.21580/jsw.2020.4.1.4544

Sutrisno, E. (2019). Aktualisasi Moderasi Beragama di Lembaga Pendidikan. *Jurnal Bimas Islam*, *12*(2), 323–348. https://doi.org/10.37302/jbi.v12i2.113

Untoro, V., & Putri, M. A. (2019). Status Identitas dan Toleransi Beragama pada Remaja. *Jurnal Psikologi Teori Dan Terapan*, *10*(1), 46. https://doi.org/10.26740/jptt.v10n1.p46-59

Van Der Linden, W. J. (2010). Item Response Theory. *International Encyclopedia of Education*, 81–88. https://doi.org/10.1016/B978-0-08-044894-7.00250-5

Willems, R. M., & Peelen, M. V. (2021). How context changes the neural basis of perception and language. *IScience*, *24*(5), 102392. https://doi.org/10.1016/J.ISCI.2021.102392