JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v13i2.39054
http://journal.uinjkt.ac.id/index.php/jp3i

Examining the psychometric properties of the Prosocial Behavior Scale Using Indonesian Pesantren (Islamic Boarding Education System) Sample

Wahyu Syahputra[1], Ika Widhiastuti[2], Baydhowi[1], Saiful Falah[3], Devie Yundianto[3], Moondore Madalina Ali[4]
[1]Department of Psychology, Faculty of Humanities, Bina Nusantara University
[2]Faculty of Psychology, State of Islamic University Syarif Hidayatullah
[3]Faculty of Tarbiyah and Teacher Science, Ummul Quro Al Islami Institute
[4]Department of Psychology, Faculty of Social, Nahdlatul Ulama Indonesia University

E-mail: Wahyu.syahputra@binus.ac.id

Abstract

In the context of behavior, prosocial behavior is needed to maintain individual relationships with others. This behavior has existed among Islamic college students known as *Santri*. *Santri*, as an adolescent, has the concept of establishing relationships through prosocial behavior. Therefore, a measuring instrument is needed to determine *Santri*'s prosocial behavior. The Prosociality Scale is one of the popular instruments in measuring prosocial behavior, and it is essential to adapt and validate the instrument with the characteristics of the Indonesian sample, especially *Santri*. This study presents evidence of validity and reliability using the Rating Scale Rasch Measurement Model. With a sample of 742, 424 (57.1%) were female and 318 (42.9%) were male. This instrument has met the criteria of good psychometrics according to the Rasch model regarding reliability (0.82 - 0.99) and validity and fulfills the need for measurement invariance.

Keywords: Adolescents, Prosocial Behavior, Rasch Model, Santri, Validation

Abstrak

Dalam konteks perilaku, perilaku prososial diperlukan untuk menjaga hubungan individu dengan orang lain. Perilaku ini telah ada di kalangan peserta didik yang menempuh pendidikan dan mendalami ilmu agama Islam di Pesantren yang dikenal dengan sebutan Santri. Santri sebagai seorang remaja memiliki konsep dalam menjalin hubungan melalui perilaku prososial. Oleh karena itu, diperlukan alat ukur untuk mengetahui perilaku prososial Santri. Skala Prososial merupakan salah satu instrumen yang populer dalam mengukur perilaku prososial, dan sangat penting untuk melakukan adaptasi dan validasi instrumen dengan karakteristik sampel Indonesia, khususnya Santri. Penelitian ini menyajikan bukti validitas dan reliabilitas dengan menggunakan Rating Scale Rasch Measurement Model. Dengan sampel sebanyak 742, dimana 424 (57,1%) adalah perempuan dan 318 (42,9%) adalah laki-laki. Instrumen ini telah memenuhi kriteria psikometrik yang baik menurut model Rasch dari segi reliabilitas (0.82 - 0.99) dan validitas serta memenuhi kebutuhan invariansi pengukuran.

***Kata Kunci**: Remaja, Perilaku Prososial, Santri, Validasi*

Introduction

In Indonesia, Islam has been acculturated and transformed into a majority religion that prioritises interfaith harmony (Anwar, 2018) and respects differences in ethnicity, culture, and language to maintain an intact social order (Maate, 2017). With adherents reaching 229.62 million people, or around 87.2% of Indonesia's total population (Mastuki, 2020), Islam provides guidance in education that prioritises religious values. This education is manifested in Pesantren, the oldest educational institution in Indonesia (Marzuki et al., 2020), which was adopted from the Hindu Buddhist educational tradition. Those who study in pesantren are called *Santri*. The root of the word pesantren itself comes from the word *Santri*, defined as a student, or earlier, Shastri, someone with expertise in explaining the holy books of Hinduism (Islam & Aziz, 2020).

Santri cannot be separated from the pesantren tradition and is the most essential part of the survival of the pesantren education pattern today (Dhofier, 2011). The reason is that their relationship and emotional connection last longer after completing their Islamic education in Pesantren. As a result, pesantren, which encourages *Santri* to continue learning and provide benefits to society, will continue to be a reference for *Santri's* behavior throughout their lives (Baso, 2012).

In terms of psychological characteristics, *Santri* shows a higher meaning of life, life optimism, and prosocial behavior than those who are not *Santri*. For example, *Santri* can adapt to any condition, thinking that the suffering that comes to them is considered wisdom to continue to be optimistic about life (Nashori, 2011). Another example of *Santri*'s character cannot be separated from their participation in the independence of the Republic of Indonesia under the formation of Laskar Hizbullah in 1943. Then, Laskar Hizbullah and Laskar Pembela Tanah Air became the forerunners of BKR and TKR (later becoming the first military institutions in Indonesia (Zuhri, 2013)). The attitude of defending the homeland and sacrificing and prioritising the community's interests is one of the characteristics of social behavior (Wittek & Bekkers, 2015).

Prosocial behavior in *Santri* cannot be separated from Islamic principles. Islam as a religion emphasises the teaching to help and assist others without expecting anything in return (Abdel-Khalek, 2013); more generally, Islam teaches compassion (Hanafi, 2001). Therefore, *Santri's* behavior is very close to social behavior (Iffan, 2019). The concept of prosocial behavior covers a wide range of things, not limited to helping, sharing, entertaining, contributing, or offering services to someone else, but all things intended to benefit others. The study of prosocial behavior examines the factors that contribute to the behavior and its impact (Dovidio & Banfield, 2015).

There is currently a notable gap in research measuring prosocial behavior among Santri, the students of pesantren (religious educational institutions), both in Indonesia and internationally. The only study in this area, by Rosset (2016), highlights differences in *Santri's* prosocial behavior but relies solely on t-test analysis. This oversight in research is significant because prosocial behavior is a fundamental element of Santri's social and moral development (Sabiq, 2012).

As a pesantren student, Santri is expected to embody positive character traits, morals, and values. Pesantren shapes individuals with noble character and integrity (Huda et al., 2023). Prosocial behaviors—such as empathy, caring, cooperation, and helping others—are essential for fostering these qualities (Baldassarri & Abascal, 2020). Engaging in prosocial behavior helps Santri understand the importance of contributing to society and reinforces their role as active, positive members of their community. These behaviors encourage active listening, respect for others' viewpoints, and a readiness to assist when support is needed, enhancing their ability to build meaningful, supportive relationships (Barr & Higgins-D'Alessandro, 2007).

In its development, prosocial behavior in *Santri* can prepare them for a better adulthood. For example, the social and emotional skills gained through prosocial behavior will help them overcome social challenges and cooperate in various environments later in life (Akelaitis & Lisinskiene, 2018; Brownell, 2013; Kaltwasser et al., 2017). Prosocial behavior in *Santri* can also help them build understanding and tolerance for differences and diversity and promote inclusiveness in their social environment (Baldassarri & Abascal, 2020). Due to the importance of prosocial behavior in *Santri*, they are expected to get holistic self-development to grow into balanced and quality individuals.

Santri in Indonesia today are, in fact, adolescents whose development is strongly influenced by their surroundings and who try to influence their environment. Prosocial behavior in adolescents, defined as voluntary actions meant to benefit others (Eisenberg et al., 2006), has been associated with several kinds of positive results, such as strong interpersonal bonds, academic achievement, and high self-esteem (Laible et al., 2004; Padilla-Walker & Carlo, 2014; Wentzel, 1993). Few longitudinal studies have looked at changes in prosocial behavior across a broad age range in adolescence, even though prosocial development has long been studied, and general age-related increases have been reported from infancy through early adulthood (Crocetti & Rubini, 2017; Eisenberg & Fabes, 1998). Examples of these studies include Carlo (2015) and Luengo Kanacri (2013).

However, adolescence brings various behavioral, cognitive, and physical changes affecting social functioning. First, adolescents may be able to participate in a greater range of prosocial activities due to their increased physical maturity and level of autonomy (Fabes et al., 1999; Padilla-Walker & Carlo, 2014). Second, improving perspective-taking (Van der Graaff et al., 2014, for example) might help people engage in prosocial conduct by facilitating higher-order moral thinking (Blasi, 1980; Eisenberg et al., 2006). Third, as social competence rises, so does the frequency of peer interactions and interest in close, romantic relationships (Steinberg & Morris, 2001).

At this stage, measurement instruments for *Santri* as adolescents are warranted. One of the reputable instruments for measuring prosocial behavior is the APBAScale developed by Caprara et al. (2005). The instrument can certainly be translated into Indonesian. Instrument adaptation is common due to the increasing need for an accurate and reliable instrument and the unavoidable diversity of cultures, languages, and races, like the adult prosociality behavior instrument, which aims to integrate Indonesian conditions with Italian culture despite the instrument's Italian origins and development. However, the necessity for instrument adaptation cannot be guaranteed to produce accurate and reliable results. In order to create two equal instruments, the proper procedures must be followed, one of which is adhering to the International Test Commission's (ITC) recommendations for the cross-cultural translation and adaptation of psychological instruments.

In addition, the Indonesian version of the Prosociality Scale was determined to be trustworthy and to meet the goodness of fit index criteria after being modified and subjected to confirmatory factor analysis using a sample of university students (Sefianmi et al., 2023). However, we found no explanation of the method of testing the results of the adaptation of the measuring instrument. Sefianmi et al. (2023) only lists the use of the confirmatory factor analysis method without explaining the shortcomings of the method when used on categorical or ordinal data. As a result, bias will occur in the resulting score because confirmatory factor analysis treats categorical data as if it were a continuum (Cai, 2010).

One psychometric method that is qualified to test the validity and reliability of a scale or instrument is the Rasch Mode (Rasch, 1960). The advantage of this method is that individuals and items can be compared on the same straight line on the logit scale (Andrich & Marais, 2019). Thus, individuals and items can be compared starting from the difficulty level of the item and the individual's ability to

respond to the item (Hayat et al., 2021), as well as the concept of objective specification, where item calibration is not tied to the person, and the estimated person score is not tied to the item (Bond & Fox, 2015).

Thus, good methods in testing instruments and the importance of measuring prosocial behavior in *Santri* provide individual benefits and encourage the formation of a better society.

Defining and Measuring Prosocial behavior

Prosocial behavior is any action or pattern of behavior that seeks to improve the welfare of others or the general population. In positive psychology, prosocial activity is the opposite of hostile behavior (Malonda et al., 2019). This is because harmful, hostile, and self-serving behaviors distinguish aggressive conduct. In contrast, prosocial behavior comprises actions that are beneficial, cooperative, and caring towards others.

Furthermore, prosocial action is typically described as voluntary, implying that people engage in it of their own free will and without any outside pressure or force (Lam, 2012). Prosocial behavior is motivated by a genuine desire to assist others rather than self-interest, the expectation of reward, or public recognition. Although no single definition is accepted worldwide, prosocial behavior is typically understood as behavior intended to help others without expecting anything in return or for one's benefit. One can see these behaviors in various contexts, such as those of families, friendships, communities, and even society (Eisenberg et al., 2010).

One instrument that measures prosocial behavior is The Prosociality Scale developed by Caprara (2005) among an Italian sample. Each of the four dimensions of the Prosociality scale will have four items, making 16 items in this instrument. The Prosociality Scale is a widely used scale for measuring adult prosociality. The Prosociality Scale is a self-report measure that assesses an individual's prosocial behavior. Participants rate the truth of each statement on a five-point Likert scale using the following options: never/nearly never true, rarely true, sometimes true, frequently true, and almost always/always true for each of the measure's 16 items. Without concentrating on particular prosocial behavior patterns, the scale evaluates adults' general prosocial behavior. The scale's components capture prosocial behaviors, including empathy and sympathetic reactions and sharing, helping, and caring behaviors (Martí-Vilar et al., 2020).

For several reasons, The Prosociality Scale differs from other prosocial measures. The Prosociality Scale, as we know, evaluates prosocial behavior in adults. At the same time, additional measures may be customised for children or adolescents (Zhan et al., 2023). Furthermore, The Prosociality Scale evaluates prosocial behavior in general rather than emphasising specific prosocial behavior classifications (Luengo Kanacri et al., 2021; Martí-Vilar et al., 2020), and the overall Prosociality Scale has strong reliability and validity and has been used in numerous studies across various countries (Luengo Kanacri et al., 2021; Zhan et al., 2023). There is some proof that the take care of scale's usefulness varies among cultures. For instance, a research investigation conducted in Italy and Spain discovered that the Spanish version of the scale showed appropriate model fit in both the Spanish and Italian groups and the results duplicated the instrument's adequacy reported in the Italian sample (Martínez-Pampliega et al., 2018).

Given the importance of measuring prosocial attitudes in Indonesia, especially for santri and the lack of research on these variables including in the context of santri in Indonesia, this study aims to test prosocial instruments to obtain satisfactory psychometric conditions. The analysis in this study uses the Rasch model, namely the rating scale measurement model, to get an overview and evidence of the fit or

not of prosocial items that follow the specific objectivity rules of the Rasch model. In addition, this study will first check the factor structure of the prosocial instrument, to find a unidimensional or multidimensional model to better explain later in the unidimensional description of the Rasch model version.

Methods

Procedure

An online survey was conducted using a Google form, to which 742 *Santri* responded. According to APA guidelines on research ethics, the researcher adhered to ethical standards when using the non-probability sampling technique. Before the poll, all respondents were aware that their participation was voluntary. Additionally, if they believed there was an irregularity in the survey administration, respondents could withdraw at any time during the survey. Respondents received notice that any personal information they unintentionally provided during the survey would be kept private, and all identifying information that may be used to link them to particular personal data was anonymized. In order to ensure the transparency of the data in this study, the data can be accessed at this link: https://zenodo.org/doi/10.5281/zenodo.13831300 while adhering to the privacy of the respondents' identities.

Respondents

Several categories were created from the 742 *Santri* that completed the poll. 424 (57.1%) *Santri* from the gender category were female as *Santriwati*, and 318 (42.9%) were male as *Santriwan*. The mean age of the study sample was 16.96 (SD = 1.34).

Instrument

In this study, the Prosocial Scale—an instrument that measures prosociality—was the object of validation. This 16-item scale, created by Caprara (2005), includes three dimensions: helping, sharing, and caring for others. While the empathy component is still debated whether it should be included in the prosocial framework, the first three dimensions are characteristic of the measurement of prosocial behavior. People's empathic motivation or disposition correlates with their desire to act prosocially. It is also a fundamental component of such tendencies (Eisenberg & Fabes, 1998). However, the dimensions may change according to the characteristics of the sample. As in Martínez-Gregorio (2023) study with a sample of adolescents in Spain, the factor structure of Capara's prosocial scale became one-dimensional. So later, we will also test the structure using factor analysis to see a better picture of the factor structure with our sample. Regarding response patterns, the scale uses a four-point Likert model, with one point (strongly disagree) to four points (strongly agree).

Translation Procedure

The Indonesian Prosociality Scale refers to the International Test Commission guidelines regarding the test's translation and adaptation procedures. First, two researchers who have an understanding of and fluency in English and Indonesia have translated the Prosociality Scale into Indonesian. Then, the translation results were discussed through a group discussion forum to discuss translation misunderstandings, and several language modifications were made at this stage. Secondly, translations into English were returned from the result to compare with the original version. Professional linguists and psychologists do the re-translation. Finally, a discussion group forum was held to evaluate factors and items that had been translated into Indonesian. The reviewers involved, ranging from forum members to translators, are holders of doctoral degrees and professionals from several universities in

Indonesia. From the result of the translation, the Indonesian Prosociality scale has a total of 16 items, which are divided into four items to helping (items 1, 3, 6, and 7), four to sharing (items 2, 9, 11, and 14), four to caring (item 4, 10, 13, and 14), and four to empathy (item 5, 8, 12, and 16)

Statistical analysis

Confirmatory Factor Analysis

The CFA model ascertains the relationships between the observable and latent variables. One kind of structural equation model is CFA, which offers an effective way to test various hypotheses regarding several measurable variables (Flora & Curran, 2004). In this study, we want to analyze the structural validity of the prosocial behavior instrument, which consists of four dimensions: helping, sharing, taking care of others (caring), and empathy. The first test was conducted by establishing a unidimensional model, then testing with four dimensions and testing with a higher-order factor model. The analysis used Mplus 8.4 software (Muthen & Muthen, 2017), where we used the weighted least squares mean estimator and the variance corrected estimator.

This study employed a number of goodness-of-fit indices to evaluate the adequacy of the proposed models. The individual model was assessed using the chi-square test. The degree to which the model did not fit well in comparison to a perfect model was estimated using the standardized root mean squared residual (SRMR) and the root mean square error of approximation (RMSEA) (Browne & Cudeck, 1993). Greater misspecification is associated with larger values; an RMSEA value of less than 0.05 is thought to be indicative of a model that fits adequately, and smaller SRMR values are linked to models that fit better; scores of less than 0.05 are thought to be indicative of a good fit (Maydeu-Olivares & Joe, 2014). Lastly, the comparative fit index (CFI) (Bentler, 1990) was also used as an incremental fit index. The fit between the given and null models could be compared thanks to the CFI. A CFI score of 0.90 is typically indicative of an appropriate model (Hu & Bentler, 1999).

The Rasch Model

The Rasch Model (Rasch, 1960) is a model that determines the probability of the response to each item based on the individual's level of the latent construct (i.e., the *Santri*'s prosocial level) and the item's level of difficulty. The model can be described as:

$$P(X_{pi} = 1) = \frac{e^{(\beta_p - \delta_i)}}{1 + e^{(\beta_p - \delta_i)}}$$

Where β_p denotes the prosocial latent concept level and δ_i denotes the item's difficulty level. Therefore, it may be said that if students can exhibit prosocial activity, δ_i provides an estimate of the rise in prosocial behavior. Initially merely dichotomous, the Rasch Model developed into a polytomous model, which started with the Rating Scale Model (Andrich, 1978). The function of the rating scale model was as follows:

$$log\ (P_{nik}/P_{ni\ (k-1)})\ =\ B_n + D_i + F_k$$

$P_{ni\ (k-1)}$ is the probability for *n* person to select the (*k - 1*) category, and P_{nik} is the likelihood that *n* person would complete item *i* in category *k*. In the meantime, D_i denotes the item's degree of difficulty (assuming that, in the statement of agreement, the respondent's difficulty is the difficulty of agreeing with the statement), and B_n level denotes the ability of the n-th individual. The likelihood that category *k* will be chosen based on category *k - 1* is thus represented by F_k. A logit scale or log odds ratio is used to express D_i and B_n (Linacre, 2006).

Because the RSM matches the instrument with the Likert format and the identical response category for every item, it will be used in this investigation. In its analysis, RSM will identify the item's placement at the difficulty level and the person's location at the trait level (Wright & Masters, 1982). In addition, RSM still has assumptions that must be met, specifically, unidimensionality, parallel item characteristic curves, and local independence (Mair, 2018). Validity tests were carried out using Winteps 3.63 based on the Rating Scale Model.

Result and Discussion

Confirmatory Factor Analysis

The fit indices of the four-dimensional, higher-order factor and unidimensional models were compared using the entire sample. The CFA results showed that the unidimensional model reasonably fit the data [$\chi2$ (90) = 202.281, $p < 0.001$; RMSEA = 0.041 (90% CI = 0.033 -0.049), CFI = 0.941, SRMR = 0.052], compared to four dimensions model [$\chi2$ (97) = 474.157, $p < 0.001$; RMSEA = 0.072 (90% CI = 0.066 -0.079), CFI = 0.802, SRMR = 0.072] and the higher order factor model [$\chi2$ (100) = 520.836, $p < 0.001$; RMSEA = 0.075 (90% CI = 0.069 0.082), CFI = 0.779, SRMR = 0.074]. Hence, this result indicates that a unidimensional model is more appropriate for this study sample. All items have significant loadings in the range of 0.147 - 0.602, of which item 1 needs special attention due to its small factor loadings.

The Rasch Model

Dimensionality

PCAR was utilized to examine the measuring instrument's unidimensionality assumption (Chou & Wang, 2010; Smith, 2002). The prosocial instrument's measurement model turned out to be unidimensional. As a result, the analysis's findings verified that the unidimensionality requirement of the Rasch RSM had been satisfied and that more research was warranted. Based on PCAR, a test is considered to measure a dimension only if the measure's minimal variance explained is greater than 30% (Linacre, 1998). The test used here demonstrated unidimensionality, as evidenced by values greater than 38.5% (16.0 in eigenvalues unit) of the variance explained by the measure. This conclusion is consistent with the findings regarding the factor structure after the CFA analysis.

Local Independence

Local independence is the basis of the Rasch RSM. For an examinee or examinees at a certain competence level, local independence refers to the ability of one item's performance to stand alone from another (Mair, 2018). The assumption of local independence was examined using the Q3 statistic (Yen, 1984) once it was established that the unidimensionality assumption had been satisfied. No items showed signs of local dependence when analyzed using the Q3 statistic index criteria, which state that the raw residual correlation between pairs of items is never > 0.30 (Christensen et al., 2017; Das Nair et al., 2011). Items 9 and 11, at 0.23, or less than 0.30, had the highest raw residual correlations. As mentioned earlier, the research assumption of local independence has been met from the report.

Item Fit

To ascertain how well each item helps to create a single common construct as proof of scale unidimensionality, item fit metrics like infit and outfit MNSQ statistics can also be adopted. Rasch RSM requirements say that an outfit or infit MNSQ value of 1 is desirable, while values between 0.5 and 1.5 work well for measurement (Andrich & Marais, 2019; Bond & Fox, 2015). Using infit and outfit MNSQ data, it was determined that all 16 items on the prosocial instrument were within the permissible range of 0.5–1.5. Furthermore, the prosocial instrument's point measure correlation range, as presented in

Table 1, was 0.43 to 0.61, meaning that every item had changed in the same direction (Bond & Fox, 2015). This outcome supported the Rasch RSM's conclusions in this investigation. Rasch item-fit statistics generally confirmed the unidimensionality of the prosocial instrument scale.

Rating Scale Diagnostic

The step order description (see Table 2) reveals that the average participant who chooses order one on this scale is estimated to be –3.14, which means that the average ability increases as the order increases or can be interpreted from negative to positive. In addition, the threshold between points 1 and 2 is -1.82, points 2 and 3 are –0.47, and points 3 and 4 2.36. Therefore, the scale threshold increases as each scale point increases from negative to positive. There was no step disorder on this scale, indicating that the four-point scale used in the prosociality scale was functioning at the same response level as intended by the test developers. The category answer function of the prosocial instrument is represented graphically in Figure 1 by the results of the scale analysis. The graph depicts the suggested pattern, where the most likely response for each ability level at each scale competency is linked. Overall, the analysis showed that the rating scale performed as intended.

Table 1. Item Fit Measure

Item	Measure	SE	Infit	Outfit	Rpm
Item 1	1.55	0.05	1.28	1.31	0.43
Item 2	-0.05	0.06	0.77	0.76	0.60
Item 3	0.48	0.06	0.96	0.94	0.51
Item 4	1.29	0.06	1.01	1.04	0.50
Item 5	0.81	0.06	1.06	1.10	0.53
Item 6	0.67	0.06	0.92	0.94	0.57
Item 7	0.35	0.06	0.93	0.93	0.52
Item 8	0.05	0.06	0.89	0.87	0.61
Item 9	-0.46	0.06	0.94	0.91	0.52
Item 10	-1.35	0.07	1.30	1.22	0.43
Item 11	-0.93	0.07	0.91	0.84	0.53
Item 12	0.36	0.06	1.15	1.15	0.53
Item 13	-0.29	0.06	0.97	0.94	0.55
Item 14	-0.83	0.07	1.05	1.00	0.52
Item 15	-1.20	0.07	1.05	0.98	0.50
Item 16	-0.44	0.06	1.03	0.99	0.54

Note: *Rpm is a point-measure correlation*
Source: Personal data (2024)

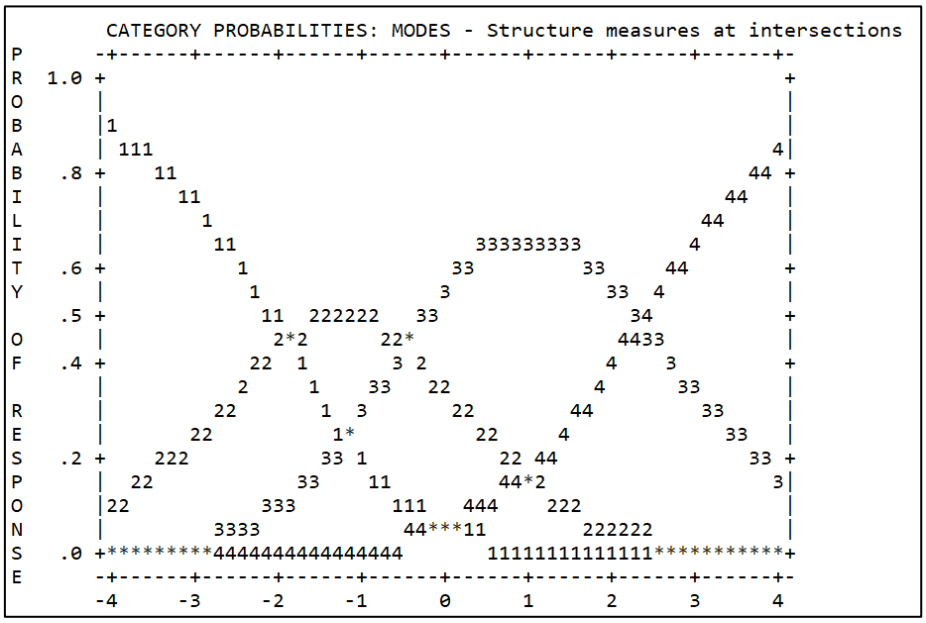
Table 2. Category Structure

Responses	Category Score	Measure	Infit MNSQ	Outfit MNSQ	Andrich threshold
Strongly Disagree	1	-3.14	1.21	1.3	-
Disagree	2	-1.21	0.95	0.95	-1.89
Agree	3	1.00	0.94	0.89	-0.47

http://journal.uinjkt.ac.id/index.php/jp3i

Strongly Agree	4	3.50	1.01	0.99	2.36

Infit MNSQ: information-weighted mean square statistic; Outfit MNSQ: outlier-sensitive mean square statistics
Source: Personal data (2024)



Source: Personal data (2024)

Figure 1. The category response curve of all Prosocial items

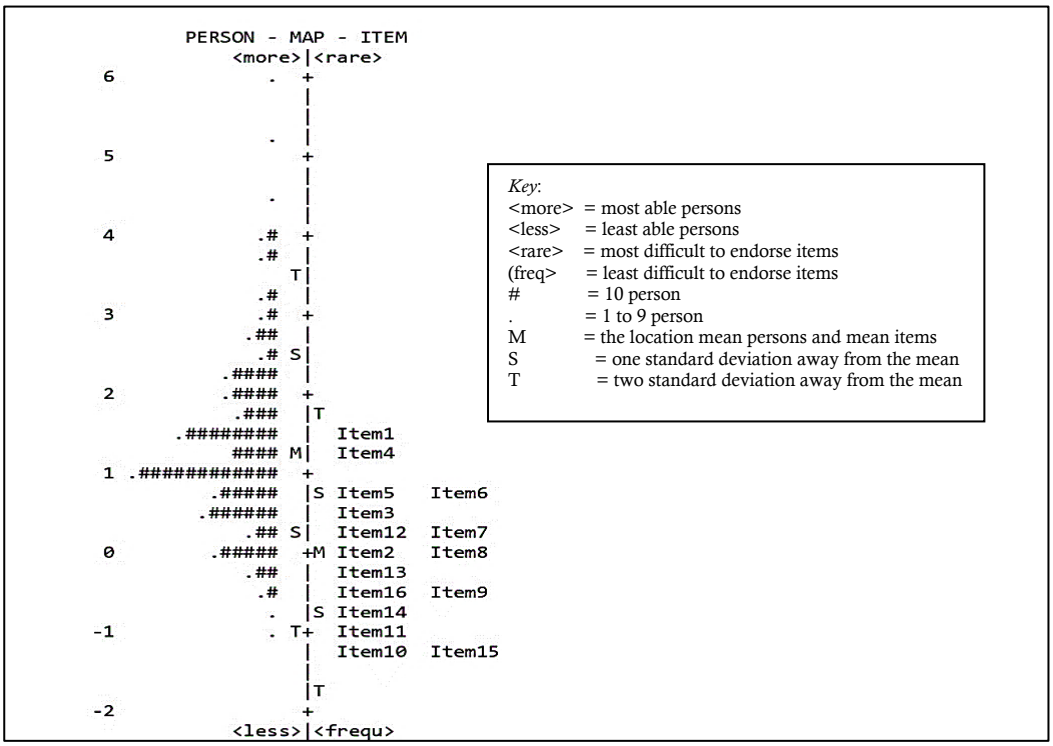
Reliability

Reliability is calculated in the Rasch RSM for both individuals and objects. For the prosocial instrument, the person separation reliability (a measure of how successfully an instrument separates people based on the variable being measured) was 0.82 (Malec et al., 2007; Wright & Masters, 1982). Thus, 2.16, expressed in standard error units, was the person separation index used to estimate the distribution of people on the measured variable. This value suggested good interpersonal separation. Excellent psychometric features were shown by the Prosocial instrument, with reliability and separation for items calculated in the same way as for individuals, coming in at 0.99 and 12.80, respectively. The provision of person reliability is above 0.80, and item reliability is above 0.90 (Malec et al., 2007). From this value, the consistency of person and item on the Indonesian Prosocial Scale is good.

Wright map

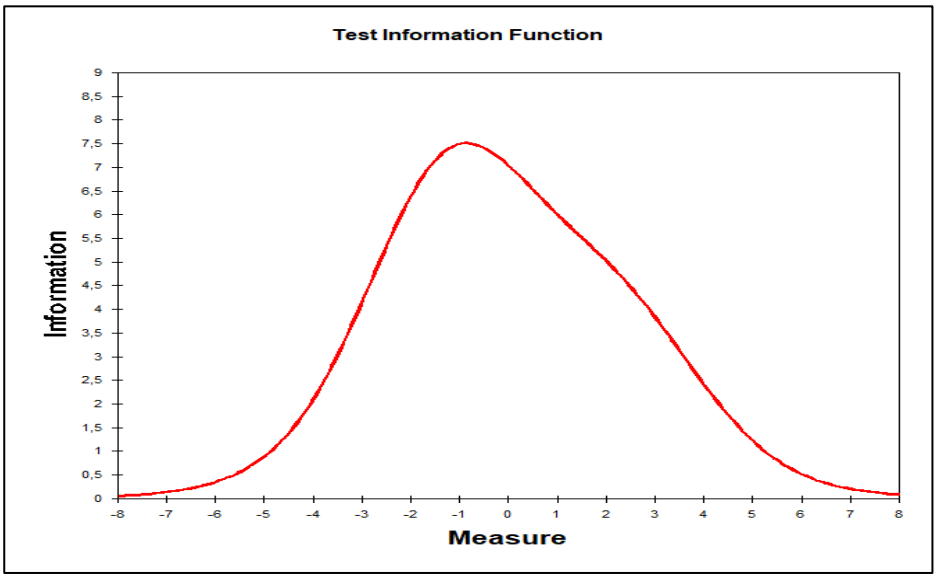
The distribution of people and items as shown by a Wright map. Information on the comparison between person and item locations is displayed in the visualization in Figure 2. Each # describes ten persons, and "." is a range of 1 to 9 persons, with the ability location on the left containing 742 participants and the difficulty location containing 16 items.

Without relying on one another's estimates, the Rasch model offers individual estimates that can be compared to the results of calibrating item difficulty in logit units. The comparison shows that the item's mean—which in the Rasch model is always 0—is lower than the individual's mean, which may signify that the person is more adept at or in agreement with the prosocial scale than the item's response difficulty. Each item has tough and easy values, and no size is lost too far compared to a meter or ruler between objects. There is also no significant gap between items in their logit location. However, the prosociality scale's 16 items fall short of the required proficiency for measures two logit or higher.



Source: Personal data (2024)

Figure 2. Item and Person Map



Source: Personal data (2024)

Figure 3. Test information function of Prosocial

Test Information Function

Finding the point on the agreement scale where a test yields the most accurate estimate of prosocial agreement is of great interest once the pertinent data for each item has been identified. The test information function (TIF) θ is only the total of the information functions for each item and Figure 3 shows the TIF for the scale. The results of the prosocial instrument showed that the relatively moderate opportunity level for agreement had the highest information function, and a significant amount of information was gleaned from the data. The latent trait level and the information derived from the measurements were relatively low at a high opportunity level for agreement. These findings demonstrated

http://journal.uinjkt.ac.id/index.php/jp3i

that the best information was obtained when the prosocial instrument was administered to those with intermediate trait levels.

Table 3. Differential Item Functioning

Item	Male		Female		DIF contrast	Rasch–Welch	
	Measure	S.E	Measure	S.E		t	p
Item 1	1.62	0.08	1.49	0.07	0.14	1.22	0.223
Item 2	-0.05	0.09	-0.05	0.08	0.00	0.00	1.000
Item 3	0.43	0.09	0.52	0.08	-0.09	-0.75	0.456
Item 4	-1.22	0.11	-1.46	0.10	0.24	1.66	0.097
Item 5	0.09	0.09	0.02	0.08	0.07	0.58	0.559
Item 6	0.67	0.09	0.67	0.08	0.00	0.00	1.000
Item 7	0.81	0.09	0.81	0.08	0.00	0.00	1.000
Item 8	-1.23	0.11	-1.20	0.09	-0.03	-0.22	0.823
Item 9	-0.49	0.10	-0.46	0.09	-0.02	-0.17	0.863
Item 10	1.29	0.08	1.29	0.07	0.00	0.00	1.000
Item 11	-0.97	0.10	-0.93	0.09	-0.03	-0.24	0.812
Item 12	-0.29	0.10	-0.29	0.08	0.00	0.00	1.000
Item 13	0.41	0.09	0.33	0.08	0.08	0.66	0.509
Item 14	-0.86	0.10	-0.83	0.09	-0.03	-0.20	0.843
Item 15	0.29	0.09	0.39	0.08	-0.10	-0.83	0.408
Item 16	-0.57	0.10	-0.34	0.09	-0.23	-1.79	0.075

Source: Personal data (2024)

Differential item functioning (DIF) testing for invariance

The Rasch model's measurement invariance creates circumstances where the calibration items are unaffected by person influence, and the expected person scores are unaffected by item influence (Wright, 2006). In essence, no single factor may lead to the specific assumption of objectivity being violated. As a result, the Rasch model, in this case, also includes DIF criteria to find the presence of objects with various functions. To ascertain the possibility of items being exposed to DIF, the reference and focal groups' scores are compared. The Mantel-Haenszel test and the Rasch-Welch t-test are the two methods for testing DIF in the Rasch model. On a scale, the latter approach is more accurate and sensitive at spotting DIF (Paek & Wilson, 2011). The logistic regression model estimates each individual or group, assuming the degree of difficulty is constant or uniform across all groups. Differences in difficulty level ratings between groups on particular items are then used as a sign of DIF, allowing some groups to receive better opportunities than others. In general, researchers can evaluate changes in answer patterns where one group gains more than another if the significance of the difference is less than 0.05. Using effect size to calculate DIF is a different strategy. According to Zieky (1993), a researcher can assume that an item indicates DIF if the contrast difference across groups on that item is more than 0.64 (Song et al., 2020). The Rasch-Welch t-test shows a value below 0.5 in probability. In this analysis, there were not any items that indicated DIF (see Table 3). However, some items could be of some concern given the differences in response patterns across genders. In item 1, "I am pleased to help my friends/colleagues in their activities," it is easier for women to agree than men. Furthermore, item 4, "I am available for volunteer activities to help those who are in need," is easier for women to agree to than men, and item 15, "I spend time with those friends who feel lonely," is more difficult for women to agree to than men.

Discussion

This study aims to test the validity of a prosociality measurement tool with a sample of *Santri* as the reference. *Santri,* as adolescents, are considered agents in modeling moderate behavior and behavior that supports social morals. So it is very important to know how the prosocial attitude of *Santri*. The Indonesian Prosociality Scale has undergone several stages of validity testing and differs from its original structure. By using the Rasch Model-based Item Response Theory, the evidence provided will be more robust and more rigorous regarding the results of the validity testing of the measuring instrument. Thus, from this study, we gained some insights.

First, The development of a four-dimensional model forming the latent prosocial variable proposed by (Caprara et al., 2005) did not support this study. After preliminary analysis of the factor structure of the prosocial scale, this research sample refers to a unidimensional model on 16 items measuring one prosocial factor. Changes in factor structure are possible in any measurement of latent variables. Especially concerning the sample characteristics used, it contains cultural, linguistic, ethnic, and other diversity (Neumann et al., 2008).

Second, the Rasch model provides evidence of the validity of the measuring instrument content using a reference to the Infit and Outfit scores. These two scores are expected to evaluate the item's validity in terms of its content. If the item exceeds the limit of the specified criteria, then the item must be removed, not the model of the calculation is changed, which makes the Rasch model a good tool in validity testing. All items on The Indonesian Prosociality Scale have passed the Infit and Outfit evaluation in the Rasch model. However, a note that certain items require review from experts, in this case, item 1 helping ("I am pleased to help my friends/colleagues in their activities"), and this is in line with the results of a low factor loading on confirmatory factor analysis that considers this item. In the Islamic boarding school environment, the words "help" and "pleased" sound strange because the culture in Islamic boarding schools makes it mandatory to help someone, especially with kindness. Adolescents often find it difficult to accept this. The attitude of helping has become a habit because of their communal life in the context of the learning place.

Third, the expansion is obtained from the description of the person and item map, which explains the comparison of the person and item on the same linear line. With logit units, The Indonesian Prosociality Scale has items that are quite easy for *Santri*, namely with the logit score of *Santri* above the logit value of the item. In this finding, the Indonesian Prosociality Scale items are unsuitable for seeing prosocial attitudes in *Santri*. The reason is that these items do not accurately describe the high prosocial attitude of *Santri,* considering that the items are quite easy to agree with. So, more precisely these items are intended to measure low individual prosocial attitudes. Developing a suitable instrument to measure prosocial attitudes in Santri is essential, with modifications to capture a range that accurately reflects Santri, who exhibits high levels of prosocial behavior. Santri is known for upholding strong religious norms, as religion emphasizes moral attitudes that support societal well-being. Moreover, the educational structure in Pesantren fosters a commitment to social responsibility, encouraging Santri to prioritize communal interests, set aside personal gains, and engage in selfless acts of support (Nahdiyah, 2018).

Fourth, the Indonesian Prosociality Scale has a good function of response order. There were no disordered step responses on the scale. This indicates something normal, and the scale's function corresponds to the test's adaptation. With the evidence of the absence of step disorder, it shows that the pattern of the respondents' answers is normal. Moreover, conversely, if step disorder occurs, then it should be noted that there is an indication that the item has been misunderstood to mean the opposite of the latent variable.

Fifth, we also found the invariance of the measuring instrument, which refers to the balance of items in measuring prosocial. In this study, gender is the object to see the fulfillment of invariance on the scale. On the invariance measurement, it can be concluded that all items are not exposed to gender bias. At least some items are easier on the other gender, for example, more favorable to women or men. However, it does not have a serious impact that causes DIF. Sometimes, in terms of gender, for example, women, we can give a picture that women have higher emotional support for colleagues than men and share behavior above men (Nielson et al., 2017), so women can answer the item more easily. Moreover, the conditions of pesantren that have nuances of independence and long-distance relationships with parents can cause a lack of feelings of affection. Women will seek friends who can relieve their problems by prioritizing mutual sharing and helping each other (Martínez-Gregorio et al., 2023).

Conclusion

The Indonesian Prosociality Scale has been psychometrically validated on a sample of *Santri*. It has good validity, although, in this study, there were changes to the factor structure that did not follow the original scale. This study illustrates that the *Santri* in the sample already have high prosocial behavior. This is possible due to the conditions of pesantren education that teach *Santri* to be kind in their social environment. As for the adaptation carried out with prosocial instruments developed outside Indonesia has yet to provide a good range in seeing the overall picture of students' prosocial behavior.

Then, we also believe that there are some limitations in this study. First, this study does not ensure external validity due to data limitations. Although it has been passed with several validity tests, external validity is considered by the researcher to be important to be carried out in future studies. External validity can be done by examining the relationship of this measuring instrument with other measuring instruments with the same concept by examining the relationship between variables that measure prosocial behavior.

Second, in this study, DIF analysis was limited to gender, although it is known that many demographic variables need to be tested and interpreted more deeply. DIF testing is important to ensure that the measuring instrument meets invariance needs.

Third, adding large-scale samples needs to be done to see the difference in the possibility of changes in the dimensional structure. Also, non-*Santri* samples are important to be re-tested and compared, given the large differences in the lifestyles of *Santri* and non-*Santri*. The Indonesian Prosociality Scale has several research constraints beyond this study's findings. Future researchers must revisit these limitations and apply them to the prosocial behavior of *Santri*.

Acknowledgment

The authors would like to thank Bina Nusantara University for funding the publication of this research. Our appreciation also extends to all who have contributed either directly or indirectly to this research.

Conflict of interest

The authors declare no competing interests.

Author's Contributions

Wahyu Syahputra research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, final approval. **Ika Widhiastuti** collection and/or assembly of data,

interpretation, writing the article. **Baydhowi** supervision, writing the article, final approval of the article. **Saiful Falah** supervision, writing the article, final approval of the article. **Devie Yundianto** data analysis and interpretation, writing the article. **Moondore Madalina Ali** supervision, writing the article, final approval of the article.

Reference

Abdel-Khalek, A. M. (2013). Personality dimensions and religiosity among Kuwaiti Muslim college students. *Personality and Individual Differences*, *54*(2), 149–152. https://doi.org/https://doi.org/10.1016/j.paid.2012.08.004

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://cir.nii.ac.jp/crid/1130000794987222016.bib?lang=en

Akelaitis, A. V., & Lisinskiene, A. R. (2018). Social emotional skills and prosocial behavior among 15-16-year-old adolescent. *European Journal of Contemporary Education*, *7*(1), 21–28. https://doi.org/10.13187/EJCED.2018.1.21

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Andrich, D., & Marais, I. (2019). A Course in Rasch Measurement Theory. In *Springer Nature*. https://doi.org/10.1016/b978-008043348-6/50005-x

Anwar, C. (2018). Islam Dan Kebhinekaan di Indonesia: Peran Agama Dalam Merawat Perbedaan. *Zawiyah: Jurnal Pemikiran Islam*, *4*(2), 1–18. https://doi.org/10.31332/ZJPI.V4I2.1074

Aryadoust, V., Ng, L. Y., & Sayama, H. (2020). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6–40. https://doi.org/10.1177/0265532220927487

Badenes-Ribera, L., Duro-García, C., López-Ibáñez, C., Martí-Vilar, M., & Sánchez-Meca, J. (2022). The Adult Prosocialness Behavior Scale: A reliability generalization meta-analysis. *International Journal of Behavioral Development*, *47*(1), 59–71. https://doi.org/10.1177/01650254221128280

Baldassarri, D., & Abascal, M. (2020). Diversity and prosocial behavior. *Science*, *369*(6508), 1183–1187. https://doi.org/10.1126/science.abb2432

Barr, J. J., & Higgins-D'Alessandro, A. (2007). Adolescent Empathy and Prosocial Behavior in the Multidimensional Context of School Culture. *The Journal of Genetic Psychology*, *168*(3), 231–250. https://doi.org/10.3200/GNTP.168.3.231-250

Baso, A. (2012). *Pesantren Studies 2a*. Pustaka Afid.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, *88*(1), 1. https://doi.org/10.1037/0033-2909.88.1.1

Bond, T., & Fox, C. (2015). *Applying the Rasch Model. In Applying the Rasch model: Fundamental Measurement in the Human Sciences*. https://doi.org/10.1017/CBO9781107415324.004

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In *Rasch Analysis in the Human Sciences*. https://doi.org/10.1007/978-94-007-6857-4

Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://econpapers.repec.org/RePEc:sae:somere:v:21:y:1992:i:2:p:230-258

Brownell, C. A. (2013). Early Development of Prosocial Behavior: Current Perspectives. *Infancy*, *18*(1), 1–9. https://doi.org/https://doi.org/10.1111/infa.12004

Cai, L. (2010). High-dimensional Exploratory Item Factor Analysis by A Metropolis–Hastings Robbins–Monro Algorithm. *Psychometrika*, *75*(1), 33–57. https://doi.org/10.1007/s11336-009-9136-x

Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A New Scale for Measuring Adults'

Prosocialness. *European Journal of Psychological Assessment*, *21*(2), 77–89. https://doi.org/10.1027/1015-5759.21.2.77

Carlo, G., Padilla-Walker, L. M., & Nielson, M. G. (2015). Longitudinal bidirectional relations between adolescents' sympathy and prosocial behavior. *Developmental Psychology*, *51*(12), 1771. https://doi.org/10.1037/dev0000056

Crocetti, E., & Rubini, M. (2017). Communicating Personal and Social Identity in Adolescence. *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.482

Davis, D. R., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) adult playfulness scale. *International Journal of Educational Research Open*, *2*, 100054. https://doi.org/https://doi.org/10.1016/j.ijedro.2021.100054

de Ayala, R. J. (2009). The Theory and Practice of Item Response Theory. In *International Journal of Testing*. https://doi.org/10.1080/15305058.2011.556771

Dhofier, Z. (2011). *Tradisi Pesantren : Studi Pandangan Hidup Kyai dan Visinya Mengenai Masa Depan Indonesia* (8th ed.). Pustaka LP3ES.

Dovidio, J. F., & Banfield, J. C. (2015). *Prosocial Behavior and Empathy* (J. D. B. T.-I. E. of the S. & B. S. (Second E. Wright (ed.); pp. 216–220). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-097086-8.24024-5

Eisenberg, N., Carlo, G., Murphy, B., & Van Court, P. (1995). Prosocial development in late adolescence: a longitudinal study. *Child Development*, *66*(4), 1179–1197. https://doi.org/10.2307/1131806

Eisenberg, N., Eggum, N. D., & Di Giunta, L. (2010). Empathy-related Responding: Associations with Prosocial Behavior, Aggression, and Intergroup Relations. *Social Issues and Policy Review*, *4*(1), 143–180. https://doi.org/10.1111/j.1751-2409.2010.01020.x

Eisenberg, N., & Fabes, R. A. (1998). Prosocial Development. In *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 5th ed.* (pp. 701–778). John Wiley & Sons, Inc.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial Development. In *Handbook of child psychology: Social, emotional, and personality development, Vol. 3, 6th ed.* (pp. 646–718). John Wiley & Sons, Inc.

Fabes, R. A., Carlo, G., Kupanoff, K., & Laible, D. (1999). Early adolescence and prosocial/moral behavior I: The role of individual processes. *The Journal of Early Adolescence*, *19*(1), 5–16. https://doi.org/10.1177/0272431699019001001

Fisher, W. P. J. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transaction*, 21–1095. https://www.rasch.org/rmt/rmt211m.htm

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Hanafi, H. (2001). *Agama, Kekerasan dan Islam Kontemporer* (1st ed.). Jendela.

Hayat, B., Hidayat, R., & Putra, M. D. K. (2021). Exploring the factor structure of environmental attitudes measure in a sample of Indonesian college students. *Revista CES Psicologia*, *14*(1), 112–129. https://doi.org/10.21615/CESP.14.1.9

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huda, M. N., Duwila, M., Nurul, S., Mojokerto, I., Gowa, D., Pungging, K., & Mojokerto, K. (2023). Menantang Disintegrasi Moral di Era Revolusi Industri 4.0 : Peran Revolusioner Pondok Pesantren. *Journal of Islamic Education*, *9*(1), 1–13. https://doi.org/10.18860/JIE.V9I1.22805

Iffan, A. G. (2019). *Santri* dan Nasionalisme. *Islamic Insights Journal*, *1*(1), 41–45. https://doi.org/10.21776/UB.IIJ.2019.001.01.4

Islam, M. H., & Aziz, A. (2020). Transformation of Pesantren in Maintaining Good Character. *HUMANISTIKA : Jurnal Keislaman*, *6*(1), 35–48. https://doi.org/10.55210/humanistika.v6i1.307

Kaltwasser, L., Hildebrandt, A., Wilhelm, O., & Sommer, W. (2017). On the relationship of emotional abilities and prosocial behavior. *Evolution and Human Behavior*, *38*(3), 298–308. https://doi.org/https://doi.org/10.1016/j.evolhumbehav.2016.10.011

Laible, D. J., Carlo, G., & Roesch, S. C. (2004). Pathways to self-esteem in late adolescence: the role of parent and peer attachment, empathy, and social behaviors. *Journal of Adolescence*, *27*(6), 703–716. https://doi.org/10.1016/j.adolescence.2004.05.005

Lam, C. M. (2012). Prosocial involvement as a positive youth development construct: a conceptual review. *TheScientificWorldJournal*, *2012*, 769158. https://doi.org/10.1100/2012/769158

Linacre, J. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Meas Trans*, *16*.

Linacre, J. (2006). A User's Guide to WINSTEPS/MINISTEP Rasch-Model Computer Programs (3.91.0). In *Chicago*. http://www.winsteps.com/a/Winsteps-ManualPDF.zip

Luengo Kanacri, B. P., Eisenberg, N., Tramontano, C., Zuffiano, A., Caprara, M. G., Regner, E., Zhu, L., Pastorelli, C., & Caprara, G. V. (2021). Measuring Prosocial Behaviors: Psychometric Properties and Cross-National Validation of the Prosociality Scale in Five Countries. *Frontiers in Psychology*, *12*, 693174. https://doi.org/10.3389/fpsyg.2021.693174

Luengo Kanacri, B. P., Pastorelli, C., Eisenberg, N., Zuffianò, A., & Caprara, G. V. (2013). The development of prosociality from adolescence to early adulthood: The role of effortful control. *Journal of Personality*, *81*(3), 302–312.

Maate, S. (2017). Peran Islam dalam Dinamika Sosial, Budaya, dan Ekonomi di Indonesia. *SOSIORELIGIUS: JURNAL ILMIAH SOSIOLOGI AGAMA*, *2*(1). https://doi.org/10.24252/SOSIORELIGIUS.V2I1.5999

Mair, P. (2018). *Modern Psychometrics with R*. Springer International. https://doi.org/https://doi.org/10.1007/978-3-319-93177-7

Malec, J. F., Torsher, L. C., Dunn, W. F., Wiegmann, D. A., Arnold, J. J., Brown, D. A., & Phatak, V. (2007). The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simulation in Healthcare : Journal of the Society for Simulation in Healthcare*, *2*(1), 4–10. https://doi.org/10.1097/SIH.0b013e31802b68ee

Malonda, E., Llorca, A., Mesurado, B., Samper, P., & Mestre, M. V. (2019). Parents or Peers? Predictors of Prosocial Behavior and Aggression: A Longitudinal Study. *Frontiers in Psychology*, *10*, 2379. https://doi.org/10.3389/fpsyg.2019.02379

Martí-Vilar, M., Merino-Soto, C., & Rodriguez, L. M. (2020). Measurement Invariance of the Prosocial Behavior Scale in Three Hispanic Countries (Argentina, Spain, and Peru). *Frontiers in Psychology*, *11*, 29. https://doi.org/10.3389/fpsyg.2020.00029

Martínez-Gregorio, S., Tomás, J. M., & Oliver, A. (2023). A Psychometric Study of the Prosocial Behavior Scale: Differential Item Functioning by Gender. In *Behavioral Sciences* (Vol. 13, Issue 3). https://doi.org/10.3390/bs13030259

Martínez-Pampliega, A., Cormenzana, S., Sanz, M., Barni, D., Simon, J., Alomar, E., & Pérez, C. (2018). Metric Goodness of the Adult Prosocialness Scale. Comparative Study of Italy and Spain. *The Spanish Journal of Psychology*, *21*, E29. https://doi.org/10.1017/sjp.2018.26

Marzuki, M., Miftahuddin, M., & Murdiono, M. (2020). Multicultural Education in Salaf Pesantren and Prevention of Religious Radicalism in Indonesia. *Jurnal Cakrawala Pendidikan*, *39*(1), 12–25. https://doi.org/10.21831/cp.v39i1.22900

Mastuki. (2020). *Menjadi Muslim, Menjadi Indonesia (Kilas Balik Indonesia Menjadi Bangsa Muslim Terbesar)*.

https://kemenag.go.id/opini/menjadi-muslim-menjadi-indonesia-kilas-balik-indonesia-menjadi-bangsa-muslim-terbesar-03w0yt

Messick, S. (1989). Validity. In *Educational measurement, 3rd ed.* (pp. 13–103). American Council on Education.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Muthen, L. K., & Muthen, B. (2017). MPlus, Statistical Analysis With Latent Variables user's guide. In *Muthén & Muthén* (8th ed.). http://www.ncbi.nlm.nih.gov/pubmed/22210541

Nahdiyah, H. (2018). Menggugah Keberadaan Pesantren Untuk Membangun Nalar Pendidikan Berbasis Local Wisdom: Suatu Tela'ah Eksploratif dalam Perspektif Budaya. *At-Turas*, *4*(2), 208–226. https://doi.org/10.33650/at-turas.v4i2.333

Nashori, F. (2011). Kekuatan Karakter *Santri*. *Millah: Journal of Religious Studies*, *11*(1 SE-Articles), 203–219. https://doi.org/10.20885/millah.vol11.iss1.art10

Nielson, M. G., Padilla-Walker, L., & Holmes, E. K. (2017). How do men and women help? Validation of a multidimensional measure of prosocial behavior. *Journal of Adolescence*, *56*, 91–106. https://doi.org/10.1016/j.adolescence.2017.02.006

Orcan, F. (2018). Exploratory and Confirmatory Factor analysis: Which one to use first? *Egit. Psikol. Olcme Deger. Derg.*, *9*(4), 414–421. https://doi.org/10.21031/epod.394323

Padilla-Walker, L. M., & Carlo, G. (2014). The study of prosocial behavior. *Prosocial Development: A Multidimensional Approach*, *3*.

Paek, I., & Wilson, M. (2011). Formulating the Rasch Differential Item Functioning Model Under the Marginal Maximum Likelihood Estimation Context and Its Comparison With Mantel–Haenszel Procedure in Short Test and Small Sample Conditions. *Educational and Psychological Measurement*, *71*(6), 1023–1046. https://doi.org/10.1177/0013164411400734

Rasch, G. (Georg). (1960). *Probabilistic models for some intelligence and attainment tests*. 199.

Rosset, D. (2016). *Perbedaan Perilaku Prososial Santriwati Antara Pesantren Modern Dan Tradisional* [Muhammadiyah Malang University]. https://eprints.umm.ac.id/34407/1/jiptummpp-gdl-debrinaros-44062-1-debrina-t.pdf

Sabiq, Z. (2012). Kecerderdasan Emosi, Kecerdasan Spiritual dan Perilaku Prososial *Santri* Pondok Pesantren Nasyrul Ulum Pamekasan. *Persona:Jurnal Psikologi Indonesia*, *1*(2). https://doi.org/10.30996/PERSONA.V1I2.21

Sefianmi, D., Purnama, C. Y., & Setyadiredja, A. N. (2023). Adaptasi Skala Prosocialness for Adult Versi Bahasa Indonesia. *Jurnal Psikologi*. https://doi.org/http://dx.doi.org/10.35760/psi.2023.v16i1.7284

Song, H.-D., Hong, A. J., & Jo, Y. (2020). Psychometric Investigation of the Utrecht Work Engagement Scale-17 Using the Rasch Measurement Model. *Psychological Reports*, *124*(3), 1384–1411. https://doi.org/10.1177/0033294120922494

Steinberg, L., & Morris, A. S. (2001). Adolescent Development. *Journal of Cognitive Education and Psychology*, *2*(1), 55–87. https://doi.org/10.1891/194589501787383444

Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W. (2014). Perspective taking and empathic concern in adolescence: gender differences in developmental changes.

Developmental Psychology, *50*(3), 881.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357. https://doi.org/10.1037/0022-0663.85.2.357

Wittek, R., & Bekkers, R. (2015). *Altruism and Prosocial Behavior, Sociology of* (J. D. B. T.-I. E. of the S. & B. S. (Second E. Wright (ed.); pp. 579–583). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-097086-8.32158-4

Wolfe, E. W., & Smith, E. V. J. (2007). Instrument development tools and activities for measure validation using Rasch models: part II--validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

Wright, B. (2006). Local Dependency, Correlations and Principal Components. *Rasch Measurement Transactions*, 509–511. https://www.rasch.org/rmt/rmt103b.htm

Wright, B., & Masters, G. (1982). *Rating Scale Analysis*. MESA Press.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zhan, Q., Wang, S., Li, C., Li, M., Liu, D., Peng, W., Song, F., Shi, T., & Li, Y. (2023). Revision and validation of the prosocialness scale for adults (PSA) among chinese college students. *BMC Psychology*, *11*(1), 124. https://doi.org/10.1186/s40359-023-01124-3

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In *Differential item functioning.* (pp. 337–347). Lawrence Erlbaum Associates, Inc. https://doi.org/10.1075/z.62.13kok

Zuhri, S. (2013). *Berangkat dari Pesantren* (1st ed.). LKiS Yogyakarta.

Appendix

Instrument (Original Version)

The following statements describe a large number of common situations. There are no right or wrong answers; the best answer is the immediate, spontaneous one. Read each phrase carefully and fill in the number that reflects your first reaction.

1. I am pleased to help my friends/colleagues in their activities.
2. I share the things that I have with my friends.
3. I try to help others.
4. I am available for volunteer activities to help those who are in need.
5. I am empathic with those who are in need.
6. I help immediately those who are in need.
7. I do what I can to help others avoid getting into trouble.
8. I intensely feel what others feel
9. I am willing to make my knowledge and abilities available to others
10. I try to console those who are sad
11. I easily lend money or other things

12. I easily put myself in the shoes of those who are in discomfort
13. I try to be close to and take care of those who are in need
14. I easily share with friends any good opportunity that comes to me
15. I spend time with those friends who feel lonely
16. I immediately sense my friends' discomfort even when it is not directly communicated to me.

Source: Caprara et al (2005)

Instrument (Indonesian Version)

Pernyataan berikut ini menggambarkan beberapa situasi tertentu. Tidak ada jawaban yang benar atau salah; jawaban terbaik adalah jawaban yang langsung dan spontan. Bacalah setiap kalimat dengan seksama dan pilihkah satu respon anda antara *sangat tidak setuju, tidak setuju, setuju dan sangat setuju* yang pada reaksi pertama anda.

1. Saya senang membantu teman/kolega saya dalam berbagai aktivitas mereka.
2. Saya berbagi sesuatu yang saya miliki dengan teman-teman saya
3. Saya berusaha membantu orang lain
4. Saya bersedia menjadi sukarelawan untuk membantu mereka yang membutuhkan
5. Saya berempati terhadap mereka yang membutuhkan
6. Saya segera membantu mereka yang membutuhkan
7. Saya melakukan apapun yang saya bisa untuk membantu orang lain agar terhindar dari masalah
8. Saya sangat dapat merasakan apa yang orang lain rasakan
9. Saya bersedia memberikan pengetahuan dan keterampilan saya kepada orang lain
10. Saya berusaha menghibur mereka yang sedang bersedih
11. Saya mudah meminjamkan uang atau barang-barang lainnya
12. Saya mudah merasakan diri saya pada posisi mereka yang sedang mengalami kesulitan
13. Saya berusaha untuk dekat dan peduli dengan mereka yang membutuhkan
14. Saya mudah berbagi dengan teman setiap mendapatkan rejeki
15. Saya menghabiskan waktu dengan teman-teman yang merasa sedih
16. Saya dapat merasakan ketidaknyamanan teman saya kepada saya meskipun dia tidak mengatakannya

Source: Personal data (2024)