

RASCH Model Application for Validation of Measurement Instruments of Student Nationalism

Andri Syawaludin*, Yetti Supriyati and Wardani Rahayu**

State University of Jakarta, Indonesia

Abstract: The aim of this study is to validate for student nationalism instruments using the Rasch Model. The reason for using the Rasch Model is because it is able to predict missing data, and verify whether the developed instrument produces expected pattern or not; calibration is done simultaneously on three things, namely the measurement scale, respondent (person), and item. The development of instruments of nationalism measurement is in order to answer and strengthen citizenship education; citizenship education is expected to restore personal identity as a nation of Indonesia and strengthen the sense of nationalism and stem the negative flow of globalization. Theoretical validation data involves 5 experts and 20 panellists; the results of quantitative theoretical validation state that all items have the reliability value of ≥ 0.70 . Empirical validation is collected from 420 public and private universities students in Samarinda, East Kalimantan, Indonesia. The results of data analysis show a good instrument reliability index ($\alpha = 0.89$), reliability of respondents ($\alpha = 0.86$), and reliability of items ($\alpha = 0.99$), and model match test (Principal component analysis) (34, 5%) has met the minimum requirement of 20%. In this study, the final results of student nationalism instruments produced 35 items that are valid and reliable.

Keywords: Rasch Model, Measurement Instruments, Measurement Instruments of Student Nationalism

Introduction

The 21st century offers limitless life, globalization, internationalization, and burst of information and technology. Thus, students are expected to excel academically and master 21st century skills to be able to face challenges (Bybee, McCrae, & Laurie, 2009). It is no possible to ignore the phenomenon of globalization in the life of a nation and state. Kunaifi (2012) argues that the development of globalization has the characteristics of: first, changes on space and time concept; second, growth of international commerce; third, increase of cultural interaction; fourth, increase of common problems.

The positive impacts of globalization are shown on the ease of information and goods flow among different countries and areas. The ease of information and goods boosts growth of science and technology. However, globalization also negatively affects the life of individuals in relation to their nationalism and statesmanship. Such negative effect is shown on the decrease of national identity. Furthermore, Kunaifi explains that globalization has changed Indonesian youngsters to favor new culture offered by the culture agent outside the school than the Indonesian culture taught at school. Thus, there is conflict of identity among the youth. The fact is strengthened by the survey results conducted by *Developing Countries Studies Center* (DCSC) Indonesia on the spirit of nationalism in commemorating Youth Pledge Day on 28 October and Heroes' Day on 10 November. The results reveal that 83.3% of respondents admit that they are either very proud or proud to be Indonesians. Meanwhile, 5.5% state that they are not proud; the rest 11.2% answer that they are not sure. However, compared to the results of similar survey in 2010 released by *Lingkar Survei Indonesia* (LSI), the nationalism spirit of Indonesians shows a decrease. The findings of LSI show that 92.1% respondents are very proud and quite proud to be Indonesians, while those who are less proud and not proud at all are as many as 4.2% and only 3.7% of them are unsure (Kunaifi & Puspita, 2012).

Husnul Wafa and Agus Satmoko (2017) opine that the decrease of nationalism among the youth is caused by disorientation, dislocation, and the tendency to prioritize personal and group needs in the name of public. For

Indonesians, such problem needs to be solved thoroughly as soon as possible considering the diverse geographical and socio-cultural characteristics of Indonesia. Through civic education, students are expected to be able to develop a high sense of nationalism. It is strengthened by research on the relation of civic education subject and increase on national concept and nationalism spirit of students. In this era of globalization, civic education is the answer to build youth with an understanding of national concept as well as the spirit of nationalism in order to solve complex nation's problems (Sofyan & Sundawa, 2015).

The development of instruments to measure students' nationalism aims to answer and strengthen the notion that through civic education we are able to answer and solve the negative effects of globalization. Eventually, it is possible to rebuild the national identity of Indonesia and strengthen the sense of nationalism. It is not quite possible to measure nationalism as it is intangible. Thus, it is necessary to build instruments or tools to collect such data.

Another main point of this study is data analysis technique used in the instrument development study through modern approach (*Item Response Theory/Rasch Model*) developed first in 1950s by Dr. Georg Rasch. The principle of Rasch model is the development of same interval of measurement, i.e. concurrent use of person score and item score to estimate true score showing the level of individual capability and the level of item difficulty (Sumintono & Widhiarso, 2015). It is different when the instrument development implemented the Classical Test Theory (CTT) in which true score is based on additive model, i.e. observed score is the sum of a true score and an error score (Allen & Yen, 1979)

There are some assumptions on classic theory. The first assumption is that an error score has no interaction with true score. The second is that an error score has no correlation to the true score and error score of test-taker's other tests. The third assumption is that the average of the error is equal to zero. The other criteria to measure test quality are difficulty index and discriminatory power. Rasch model analysis has more benefits, such as the ability to predict missing data, verify whether the developed instrument produce expected pattern or not, concurrently calibrate three things i.e. measurement scale, respondent (person), and question item (item). Through thorough analysis of Rasch model, valid data and expected instrument (reliable) are obtained (Bond & Fox, 2015).

Literature Review

Understanding of Nationalism

Smith (2003) state that nationalism is defined as (1) the development process or growth of nations; (2) the sentiment or awareness of sense of belonging to one state or nation; (3) the language and symbolisms of a nation, (4) the socio-political movements for the nation; (5) the national doctrines or ideologies, either general or specific. Shafer mentions that nationalism has multi meanings. It depends on the objective and subjective condition of each nation. Thus, nationalism has these meanings (Shafer, 1955): (1) It is the love for the same motherland, race, language, and culture; therefore in this case, nationalism is similar to patriotism. (2). It is a desire of national political freedom, security, and prestige. (3) It is a mystical service to a blur or even supernatural social organism known as a nation or volk which is superior as a union than as parts. (4) It is a dogma teaching that individuals live only for the nation itself. (5) It is a doctrine stating that a nation has to be the most dominant and highest among others and thus, it has to act aggressively. Sartono Kartodirjo proposes the definition formula of *nasion* which refers to a community of life union involving different ethnics, class or social groups, beliefs, and others. They are all integrated in historical development as a union of political system based on solidarity supported by common political desire (Kartodirjo, 1999a). In his other book, Sartono Kartodirjo (1999b) states that nationalism covers the guarantee of national unity, individual and group liberty, individual equality, personality, and performance or supremacy for the future of the nation.

Anderson in Sutarjo's writing defines nation as an imagined political community – it is imagined as something inherently limited and sovereign (Adisusilo, 2009). In this case, Anderson notes that the term imagined is due to the fact that the members of the nation commonly never see each other yet at the same time in their mind, they are the members of certain community. Thus, Sutarjo, based on Anderson's writing, understands that nationalism is a living form, which dynamically develops and finds new forms in line with the development and demand of life. The theoretical framework used as the main indicators of student nationalism measurement instrument is the one proposed by Kartodirjo; that nationalism includes guarantee of national unity, individual and group liberty, individual equality, personality, and performance or supremacy for the future of the nation. These five points are the indicators of the developed instruments. The other reference is attitude; it is one's mental condition showing their evaluation (feeling/affection) on certain objects and action tendency (conation) based on their knowledge (cognition). The implementation of this reference is based on Anderson's taxonomy which is the revision of Bloom's taxonomy. There are some changes on the use of nouns to verbs; in relation to nationalism, the aspect of cognition is related to nationalism concept; the aspect of affection is related to the sense of nationalism; and the aspect of conation is related to the spirit of nationalism.

Understanding of Rasch Model in Measurement Tool Development

The classical test theory or known as true score is based on certain additive model, i.e. observed score is the sum of true score and error score (Allen & Yen, 1979). There are some assumptions on classic theory. The first assumption is that an error score has no interaction with true score. The second is that an error score has no correlation to the true score and error score of test-taker's other tests. The third assumption is that the average of the error is equal to zero. The other criterias to measure test quality are difficulty index and discriminatory power.

Classical Test Theory (CCT) has some fundamental weaknesses as many statistics implemented in CTT such as difficulty level and discriminatory power are highly dependent on the samples used in analysis. In addition, the results of the test cannot be generalized to other tests. Furthermore, it is difficult to ensure the test reliability in the context of CTT. Additionally, CTT has no base to determine a test participant's response on certain test item, and the standard error of measurement is assumed to be the same for all test participants (Hambleton & Jones, 1993; Naga, 2013, Retnawati, 2014, Haiyang, 2010).

What makes CTT different from IRT/Rasch Model in evaluating or measuring research instrument? First, the item analysis in Rasch Model is done on each item. Second, Rasch Model concurrently tests person (respondent) in which it is able to show the consistent pattern of respondents' answer, reveal which respondents who tend to agree (in behavior instrument), and identify which respondents who tend to answer recklessly. The test for research instrument can also be done in the form of multidimensional test; Rasch Model is able to detect the bias of tested items. In brief, Rasch Model fulfills the objective measurement requirements.

Georg Rasch developed one analysis model from Item Response Theory (IRT) in 1960s commonly known as 1PL (one logistic parameter). Ben Wright helps make this mathematic model popular (Linacre, 2011). By using raw data in the form of dichotomous data (in the form of true and false), Rasch formulates a model connecting students and test item (Misbach & Sumintono, 2015).

As an illustration, a student who is able to correctly answer 80% questions surely has better ability than another student who can only accurately answer 65% test items. Such data (percentage) shows that the obtained data are ordinal data showing rank and not linear (Linacre, 1999). As ordinal data have no similar interval, it needs to be changed to ratio data for the need of statistical analysis. Thus, if someone has 80% score, the odds ratio is 80:20 (meaning 80 true answers against 20 false answers), which is a more proper frequency/ratio data for measurement goal. Through ratio data, Rasch develops measurement model to determine the relation of students' ability (person ability) and item difficulty by using logarithm function to produce measurement with

the same interval. The result is a new denomination known as logit (log odds unit) showing person ability and item difficulty. Thus, based on the obtained logit score, it can be concluded that person ability in answering questions highly depends on its level of ability and test difficulty (Sumintono, 2015).

Besides dichotomous data, Rasch Model can also be used to analyze polytomous data as the one developed by Andrich, which is similarly based on two basic theorems: person ability and test difficulty. Rasch Model assumes that test difficulty is affected by respondents' answer, and that person ability is affected by test difficulty (Misbach & Sumintono, 2015).

Model Rasch Analysis produces fit statistics analysis which provides information to the researcher on whether the obtained data ideally show that a person with a higher ability gives answer of pattern in line with test difficulty. The implemented parameters are the infit and outfit of mean square and standardized values. Sumintono and Widhiarso (2013) define infit (inlier sensitive or information weighted fit) as the sensitivity of response pattern on test items of a respondent (person); meanwhile, outfit (outlier sensitive fit) measures the sensitivity of response pattern on test item with certain difficulty on respondent and vice versa.

On the analysis of instrument level done using Winsteps software application developed by Linacre, if the data are compatible with Rasch Model, the sum of mean square is 1.0 while the score of Z-standardized value is 0.0. On the other hand, in the level of each test item or respondent, the required parameters to see the compatibility and incompatibility are:

Point Measure Correlation (x) : $0.32 < x < 0.8$

Outfit Mean Square (y) : $0.5 < y < 1.5$

Outfit Z Standard (z) : $- 2.0 < z < +2.0$

In the context of item test, a misfit item is the one which is too easy (the logit score is too negative) or too difficult (the logit score is too big) based on respondents' answers; an item is also considered misfit when the score of the three criterias obtained from the software application shows that the items do not fulfill the criteria indicating that the items do not measure the expected characteristic (Misbach & Sumintono, 2015). Through Winstep software application, Rasch Model can produce analysis with reliability on instrument level (respondent and item), respondent and item validity, unidimensional instrument, test item bias detection, and accuracy on implemented response sum (Wibisono, 2015).

The draft of test plan and test item distribution to measure student nationalism is shown on table 1:

Table 1. Instrument Plan on Student Nationalism

NO	DIMENSION	INDICATOR	ITEM		TOTAL
			favorable	Unfavorable	
1	Nationalism Concept (Cognition)	Unity	3	1	4
		Liberty	3	3	6
		Equality	2	1	3
		Personality	3	1	4
		Performance	2	1	3
2	Sense of Nationalism (Affection)	Unity	2	1	3
		Liberty	1	2	3
		Equality	2	1	3
		Personality	2	2	4
		Performance	1	1	2
3	Spirit of Nationalism (Conation)	Unity	1	1	2
		Liberty	1	1	2
		Equality	2	1	3
		Personality	2	2	4
		Performance	2	1	4
	Total Number				50

Methodology

Stages of student nationalism instrument development present the description of instrument standardization process with the following results:

a) Theoretical Validation/Theoretical Trial through expert and panellist quantitative and qualitative assessment. Qualitative assessment involves 5 experts while quantitative assessment involves 20 panellists. The initial stage involves 25 people in total and is supported with FGD (Forum Group Discussion) collaborating with a *widyaiswara* (like a lecturer).

b) Empirical Validation is conducted in two phases. The first phase empirically involves 150 students as the respondents; meanwhile, the result of the study stating that 42 items in the first trial phase are valid involves 420 respondent; referring to the determination of the number of respondents or samples according to some experts such as Gable, the number of subjects or respondents is supposed to be 6 up to 10 times of the items which are going to be analysed to obtain the trial data (Gable & Wolf, 1993). Therefore, the total number of the respondents in the second phase is $42 \times 10 = 420$ students. In general, the research population is the Civic Education students. The chosen students are from state as well as private universities in Samarinda, East Borneo. The sampling technique applied in this research is multistage random sampling aimed to select the universities and determine the respondents.

c) Data Analysis

The data are analysed using Rasch model and supported with Winstep software developed by Linacre (2006). This model is able to see the interaction between respondents and items at one time. Based on this model, a score does not solely refer to raw score but based more on the logit score which reflects the selection probability of an item on a group of respondents. It is used as raw score anticipation from likert rating in the form of ordinal data which have no similarity of interval among the scores.

The use of Rasch model for polytomous data developed by Andrich is based on two basic theorems; the level of ability/agreement of an individual and the level of difficulty of the items (Linacre in Misbah & Sumintono, 2014). The psychometric set used in this study includes reliability in the level of instruments (respondents and items), validity of respondents and items, unidimensionality of instruments, bias detection of test item, and accuracy on the number of used responses.

Result and Discussion

The data are obtained from 420 student respondents; They are then tabulated using Ms.Excel software. After that, They are converted and analysed using Winstep 4.0.0 software on Windows Operating System.

The statistical summary data of the 420 respondents on the instruments of student nationalism is as follows:

Table 2. Summary Statistic

SUMMARY OF 420 MEASURED Person								
TOTAL			MODEL		INFIT		OUTFIT	
ZSTD	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	
MEAN .2	168.5	41.7	1.53	.22	1.07	-.1	1.04	-
P.SD 2.2	14.1	.6	.68	.03	.63	2.4	.60	
S.SD 2.3	14.1	.6	.68	.03	.63	2.4	.60	
MAX. 7.9	204.0	42.0	4.09	.43	3.73	7.5	3.90	
MIN. 5.5	133.0	38.0	.06	.18	.17	-5.8	.20	-
REAL RMSE .86	.26	TRUE SD	.63	SEPARATION	2.45	Person RELIABILITY		
MODEL RMSE .89	.22	TRUE SD	.64	SEPARATION	2.87	Person RELIABILITY		

| S.E. OF Person MEAN = .03
|

Person RAW SCORE-TO-MEASURE CORRELATION = .98

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .89 SEM = 4.66

SUMMARY OF 42 MEASURED Item

	TOTAL		MODEL		INFIT		OUTFIT
ZSTD	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ

MEAN	1685.2	417.4	.00	.07	1.01	-.1	1.04
------	--------	-------	-----	-----	------	-----	------

P.SD	171.5	2.0	.83	.01	.25	3.4	.28
------	-------	-----	-----	-----	-----	-----	-----

S.SD	173.6	2.0	.84	.01	.26	3.5	.28
------	-------	-----	-----	-----	-----	-----	-----

MAX.	1999.0	420.0	1.48	.11	1.79	9.3	1.92
------	--------	-------	------	-----	------	-----	------

MIN.	1324.0	413.0	-1.99	.06	.62	-6.4	.63
------	--------	-------	-------	-----	-----	------	-----

REAL RMSE	.08	TRUE SD	.83	SEPARATION	10.95	Item RELIABILITY	.99
-----------	-----	---------	-----	------------	-------	------------------	-----

MODEL RMSE	.07	TRUE SD	.83	SEPARATION	11.49	Item RELIABILITY	.99
------------	-----	---------	-----	------------	-------	------------------	-----

| S.E. OF Item MEAN = .13
|

Item RAW SCORE-TO-MEASURE CORRELATION = -.99

Global statistics: please see Table 44.

UMEAN=.0000 USCALE=1.0000

Respondent Analysis

The table shows that the participants' average logit is +1.53. It represents the average score of student nationalism. The average score, which is higher than 0.0, indicates that most participants tend to agree with the given statements in many question items.

The quality of all respondents can be seen from the score of Infit MNSQ and Outfit MNSQ which is better when it approaches the score of 1.0, and also the score of Infit ZSTD and Outfit ZSTD which must approach 0.0. In Person, the score of Infit MNSQ is 1.07 and it is 1.04 for the Outfit MNSQ score. Meanwhile, the score of Infit ZSTD is -0.1 and the obtained score for Outfit ZSTD is -0.2. Both data show that the quality of all respondents is good.

The important information is that the score of person reliability is 0.86 (very good) (Table 2), while the Person Separation score is rounded down from 2.45 to 2. The bigger the separation score, the better the quality of the instruments in relation to respondents and test items. It is because it can identify respondent and item group. A more accurate strata separation equation which can be used is:

H : Score of Strata Person

SEPARATION : score of SEPARATION for the obtained score of the respondents, The score of strata separation for person is $H = ((4 \times 2.45) + 1) / 3$, $H = 3.6$ and is rounded to 4. It means that there are four groups of respondents, in this case is the respondents' level or perception on nationalism: high, moderate, low, very low. Ravand, Hamdollah, and Firoozi (2016) state that person separation of more than 2 shows that the questionnaire is sensitive enough to distinguish between high and low proficiency test participants, thus supporting the external aspects of test validity. The result of person separation of nationalism attitudes of students was 4, including the excellent category (Ducan, Bode, Lai and Perera, 2003).

Item Analysis

The quality of all statement items can be seen based on the Infit MNSQ and Outfit MNSQ score that is close to 1.0 meaning that it is better. Besides, it is also based on the score of Infit ZSTD and Outfit ZSTD that must approach 0.0. The data in the table shows that the score of Infit MNSQ is 1.01 and 1.04 for the Outfit MNSQ. Meanwhile, it is -0.1 for the Infit ZSTD score and 0.2 for the Outfit ZSTD. This indicates that the all items are good.

To determine the static compatibility between the model and the data, mean-square fit statistics (MNSQ outfit) and z-standardize fit statistics (ZSTD outfit) can be used. MNSQ values are in the range of 0.5 to 1.5 while for ZSTD are in the range of -2 to 2 (Linacre, 2009; Anshel, Weatherby, Kang, & Watson, 2009; Gómez, Arias, Verdugo, & Navas, 2012; Abdullah & Lim, 2013 ; Perera, Sumintono, & Na, 2018).

The result of item reliability is 0.99 (categorized as very good). In addition, the score of item strata separation is $H = ((4 \times 10.95) + 1) / 3$, $H = 14.93$ which is rounded to 15. This means that there are fifteen groups of items: the most difficult items and the easiest ones (very difficult, difficult, --- very easy).

As for the Cronbach alpha score (measuring reliability: interaction between person and items as a whole), the obtained score is 0.89 showing that the level of reliability is very good (Sumintono & Widhiarso, 2014).

Based on the evaluation of psychometric property, it can be concluded that the obtained actual data has fulfilled the requirements of Rasch model so that the next analysis can be further applied.

The next analysis/step is testing the accuracy of individual items using the model. This can be seen in the item table: misfit order. The results of the analysis of 420 respondents are as follows:

Table 3: Misfit Order of 420 Respondents

ENTRY	TOTAL	TOTAL	MEASURE	MODEL	INFIT	OUTFIT	PTMEASUR	AL	EXACT	MATCH			
NUMBER	SCORE	COUNT	MEASURE	S. E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
Item													
N39	39	1324	417	1.48	.06	1.79	9.3	1.92	9.9	A .21	.49	34.5	46.1
N21	21	1727	419	-.06	.07	1.50	6.4	1.69	8.5	B .36	.41	48.9	49.9
N11	11	1938	418	-1.37	.09	1.47	5.1	1.29	2.9	C .36	.30	71.5	66.4
N30	30	1378	419	1.32	.06	1.33	4.3	1.42	5.3	D .22	.48	42.5	46.8
N32	32	1643	420	.33	.07	1.33	4.4	1.39	5.0	E .32	.43	42.6	48.1
N29	29	1876	419	-.88	.08	1.16	2.2	1.36	4.2	F .42	.35	67.5	57.4
N12	12	1601	413	.40	.07	1.33	4.2	1.34	4.4	G .44	.44	43.8	47.8
N9	9	1899	418	-1.07	.08	1.33	4.1	1.33	3.6	H .29	.33	60.5	60.4
N18	18	1709	419	.02	.07	1.24	3.3	1.30	4.1	I .42	.42	51.3	49.4
N7	7	1697	416	.02	.07	1.23	3.1	1.28	3.8	J .36	.41	49.3	49.5
N27	27	1393	419	1.27	.06	1.11	1.5	1.23	3.0	K .26	.48	49.6	46.9
N42	42	1386	419	1.30	.06	1.08	1.1	1.18	2.4	L .27	.48	47.3	46.9
N36	36	1811	419	-.49	.07	1.16	2.3	1.13	1.8	M .42	.38	58.0	52.7
N6	6	1910	418	-1.15	.09	1.13	1.7	1.11	1.3	N .35	.33	66.0	62.1
N41	41	1799	415	-.52	.07	1.12	1.8	1.05	.8	O .51	.38	57.6	52.8
N8	8	1428	418	1.14	.06	1.02	.3	1.09	1.2	P .33	.47	47.1	47.3
N23	23	1629	416	.32	.07	1.06	.9	1.09	1.2	Q .46	.43	49.8	48.1
N15	15	1500	415	.83	.06	.99	-.1	1.07	1.1	R .42	.46	52.0	47.6
N28	28	1426	414	1.10	.06	.96	-.6	1.07	1.0	S .31	.47	47.8	47.3
N3	3	1916	418	-1.19	.09	1.01	.1	1.03	.4	T .29	.32	63.6	62.8
N26	26	1795	418	-.43	.07	1.01	.1	.96	-.5	U .47	.39	61.5	52.2

N40	40	1604	419	.48	.06	.97	-.4	1.01	.2	u	.41	.44	50.8	47.6
N4	4	1973	420	-1.60	.10	.98	-.2	.98	-.2	t	.30	.28	68.6	71.7
N10	10	1734	414	-.20	.07	.95	-.7	.98	-.2	s	.44	.40	59.9	50.2
N25	25	1668	418	.19	.07	.92	-1.2	.97	-.4	r	.48	.43	50.7	48.6
N5	5	1776	417	-.35	.07	.95	-.7	.96	-.5	q	.39	.39	54.4	51.1
N34	34	1724	420	-.03	.07	.94	-.8	.93	-1.0	p	.47	.41	52.6	49.8
N1	1	1999	418	-1.99	.11	.90	-1.0	.93	-.5	o	.22	.25	78.2	79.4
N16	16	1642	413	.22	.07	.88	-1.8	.90	-1.5	n	.53	.43	53.0	48.5
N2	2	1959	420	-1.47	.09	.85	-1.8	.81	-2.1	m	.39	.29	71.9	68.7
N33	33	1772	420	-.27	.07	.83	-2.7	.81	-3.0	l	.59	.40	59.3	50.7
N22	22	1630	418	.36	.07	.79	-3.1	.82	-2.8	k	.46	.44	53.8	48.1
N20	20	1664	417	.19	.07	.78	-3.3	.80	-3.1	j	.48	.43	57.3	48.6
N13	13	1519	413	.74	.06	.77	-3.5	.79	-3.2	i	.53	.45	52.5	47.5
N14	14	1525	416	.75	.06	.74	-3.9	.78	-3.3	h	.47	.45	53.1	47.5
N38	38	1677	416	.11	.07	.77	-3.6	.76	-3.8	g	.58	.42	55.8	49.0
N19	19	1683	417	.10	.07	.74	-4.1	.75	-4.0	f	.53	.42	58.3	49.0
N37	37	1668	416	.16	.07	.75	-3.9	.75	-4.0	e	.57	.42	55.8	48.9
N35	35	1770	418	-.29	.07	.72	-4.5	.71	-4.7	d	.56	.39	62.0	50.7
N31	31	1575	419	.60	.06	.70	-4.7	.71	-4.5	c	.57	.45	53.5	47.5
N17	17	1724	416	-.11	.07	.63	-6.2	.63	-6.1	b	.55	.41	61.5	50.0
N24	24	1708	419	.02	.07	.62	-6.4	.63	-6.3	a	.54	.42	64.9	49.4
-----+-----+-----+-----+-----+-----+-----														
MEAN	1685.2	417.4	.00	.07	1.01	-.1	1.04	.2					55.7	52.2
P. SD	171.5	2.0	.83	.01	.25	3.4	.28	3.6					8.7	7.5

The criteria used is to test the misfit/outlier item by summing up the mean and standard deviation (S.D.) of INFIT MNSQ ($1.01 + 0.25 = 1.26$); INFIT MNSQ score which is higher than 1.26 is the candidate of misfit/outlier item, such as N39 (1.79), N21 (1.50), N11 (1.47), N30 (1.33), N32 (1.33), N12 (1.33), N9 (1.33).

Bias Detection

The bias item in this measurement is viewed based on two variables: respondents' gender and ages. Rasch model analysis displays detection of bias item in Differential Item Functioning/DIF. Bias can be detected through the item probability score which is under 5% (Sumintono and Widhiarso, 2014).

Table 4 below is the measurement of DIF to find out if there is item bias in particular groups (benefiting male group and harming female group). Here is the display of the data of DIF based on respondents' gender (male and female).

Table 4. DIF JENDER

DIF class specification is: DIF=\$S4w1

Person	SUMMARY DIF			BETWEEN-CLASS			Item	
	CLASSES	CHI-SQUARED	D.F.	PROB.	UNWTD	MNSQ	t=ZSTD	Number Name
2	6.1619	1	.0131	6.2611	2.2599	1	N1	
2	4.5033	1	.0338	4.5517	1.8657	2	N2	
2	11.4661	1	.0007	11.7270	3.1696	3	N3	
2	.6686	1	.4135	.6710	.2073	4	N4	
2	.0000	1	1.0000	.0099	-1.1951	5	N5	
2	6.5519	1	.0105	6.6423	2.3377	6	N6	
2	.3727	1	.5415	.3738	-.1218	7	N7	
2	.0000	1	1.0000	.0564	-.8364	8	N8	
2	1.7318	1	.1882	1.7411	.9021	9	N9	
2	.2647	1	.6069	.2655	-.2866	10	N10	
2	.4886	1	.4846	.4902	.0227	11	N11	
2	.0000	1	1.0000	.0124	-1.1589	12	N12	
2	7.5889	1	.0059	7.6785	2.5351	13	N13	
2	.8060	1	.3693	.8087	.3265	14	N14	
2	1.2982	1	.2545	1.3034	.6673	15	N15	
2	.0000	1	1.0000	.0575	-.8310	16	N16	
2	1.6036	1	.2054	1.6111	.8369	17	N17	
2	2.8653	1	.0905	2.8832	1.3693	18	N18	
2	4.2374	1	.0395	4.2715	1.7920	19	N19	
2	2.5985	1	.1070	2.6136	1.2721	20	N20	
2	3.5252	1	.0604	3.5509	1.5864	21	N21	

	2	.2180	1	.6406	.2186	-.3721	22	N22	
	2	.0000	1	1.0000	.0471	-.8841	23	N23	
	2	.0506	1	.8220	.0807	-.7333	24	N24	
	2	.3810	1	.5371	.3821	-.1106	25	N25	
	2	1.8317	1	.1759	1.8412	.9501	26	N26	
	2	3.8039	1	.0511	3.8310	1.6694	27	N27	
	2	1.7035	1	.1918	1.7112	.8874	28	N28	
	2	.0399	1	.8417	.0602	-.8184	29	N29	
	2	1.1808	1	.2772	1.1853	.5951	30	N30	
	2	.2100	1	.6468	.2105	-.3881	31	N31	
	2	3.4411	1	.0636	3.4644	1.5599	32	N32	
	2	1.3177	1	.2510	1.3234	.6791	33	N33	
	2	.4562	1	.4994	.4576	-.0152	34	N34	
	2	4.4574	1	.0348	4.4968	1.8515	35	N35	
	2	2.4223	1	.1196	2.4369	1.2047	36	N36	
	2	1.4510	1	.2284	1.4574	.7552	37	N37	
	2	2.3530	1	.1250	2.3661	1.1768	38	N38	
	2	1.0044	1	.3163	1.0081	.4771	39	N39	
	2	1.2723	1	.2593	1.2774	.6518	40	N40	
	2	.0000	1	1.0000	.0028	-1.3527	41	N41	
	2	.0000	1	1.0000	.0177	-1.0975	42	N42	

DIF is characterized by the score of probability which is under 5% (0.05) meaning that the item is bias. Referring to the table above, there are indicated bias items based on respondents' gender: **N1, N2, N3, N6, N13, N19, and N35.**

The table below is also related to DIF in which the criterion is based on respondents' ages. It is as presented as follows:

Table 5. DIF of 420 Respondents Age

DIF class specification is: DIF=\$S5W1

	Person	SUMMARY DIF			BETWEEN-CLASS		Item	
	CLASSES	CHI-SQUARED	D.F.	PROB.	UNWTD MNSQ	t=ZSTD	Number Name	

7	5.2574	6	.5111	.9091	.0300	1 N1
7	2.1464	6	.9057	.3659	-1.2873	2 N2
7	8.7714	6	.1867	1.5143	.9633	3 N3
7	5.7623	6	.4501	1.0408	.2621	4 N4
7	10.1823	6	.1170	1.8745	1.4031	5 N5
7	7.4044	6	.2849	1.2729	.6277	6 N6
7	4.5693	6	.6000	.7905	-.1992	7 N7
7	1.8992	6	.9287	.3281	-1.4199	8 N8
7	7.1743	6	.3048	1.2548	.6009	9 N9
7	4.1695	6	.6536	.7195	-.3476	10 N10
7	3.6167	6	.7283	.6207	-.5712	11 N11
7	14.6128	6	.0234	2.7436	2.2706	12 N12
7	7.0447	6	.3165	1.2323	.5672	13 N13
7	3.7364	6	.7122	.6367	-.5335	14 N14
7	9.3467	6	.1548	1.6215	1.1009	15 N15
7	9.6805	6	.1386	1.8023	1.3198	16 N16
7	4.5339	6	.6047	.7686	-.2439	17 N17
7	3.5359	6	.7391	.6858	-.4213	18 N18
7	4.1436	6	.6571	.7042	-.3809	19 N19
7	6.8523	6	.3345	1.1968	.5131	20 N20
7	6.1466	6	.4068	1.1465	.4347	21 N21
7	3.8799	6	.6928	.6572	-.4860	22 N22
7	6.1153	6	.4102	1.0463	.2714	23 N23
7	2.4602	6	.8729	.4174	-1.1204	24 N24
7	5.4033	6	.4930	.9232	.0559	25 N25
7	7.4613	6	.2801	1.4061	.8177	26 N26
7	5.1764	6	.5212	.8914	-.0029	27 N27
7	6.1774	6	.4034	1.0851	.3359	28 N28
7	5.9314	6	.4307	1.0606	.2954	29 N29
7	4.3061	6	.6352	.7360	-.3121	30 N30
7	1.9818	6	.9213	.3387	-1.3816	31 N31
7	8.1905	6	.2243	1.4532	.8819	32 N32
7	4.5186	6	.6067	.7976	-.1848	33 N33

	7	4.1612	6	.6548	.7149	-.3575	34	N34	
	7	2.7915	6	.8345	.4834	-.9257	35	N35	
	7	5.9450	6	.4292	1.0421	.2644	36	N36	
	7	6.0168	6	.4211	1.0349	.2522	37	N37	
	7	1.5041	6	.9592	.2609	-1.6836	38	N38	
	7	13.0197	6	.0426	2.3635	1.9178	39	N39	
	7	3.0141	6	.8070	.5116	-.8479	40	N40	
	7	4.0356	6	.6717	.6960	-.3988	41	N41	
	7	3.5329	6	.7395	.6043	-.6106	42	N42	

DIF is characterized by the score of probability which is under 5% (0.05) meaning that the item is bias. Based on the table above, there are indicated bias items based on respondents' age: **N12 and N39**.

Rating Scale Validity

The rating scale validity is a test used to verify the selected rating whether or not the selection rating confuses the respondents. Rasch Model Analysis provides verification process on rating assumption written in the instruments. In this instrument, 5 options of answer are given in the form of likert rating for each item. Respondents answer each given item. The answer is viewed based on the tendency on whether the answer moves to the left column (strongly disagree) or to the right (strongly agree).

Table 6. Rating Scale of 420 Respondents

```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-----
|CATEGORY  OBSERVED|OBSVD SAMPLE|INFIT  OUTFIT||  ANDRICH  |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
|-----+-----+-----++-----+-----|
| 1  1    187  1| .81  -.16|  1.70  2.49||  NONE   |( -2.66)| 1
| 2  2    536  3| .52*  .25|  1.26  1.38||  -1.01  | -1.37 | 2
| 3  3   3809 22| .67   .78|   .87   .85||  -1.45  |  -.25 | 3
| 4  4   6902 39| 1.39  1.43|   .91   .86||   .50   |  1.30 | 4
| 5  5   6097 35| 2.34  2.27|   .92   .94||  1.96  |(  3.21)| 5
|-----+-----+-----++-----+-----|
|MISSING    109  1| 1.19      |           ||           |           |
-----
```

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

Table 6 above uses Andrich Threshold criterion to examine whether or not the polytomous score used is already appropriate. The score of Andrich Threshold moving from NONE into negative and heading to positive in sequence shows that the given options are valid. On the contrary, the table does not present sequential data.

Instruments Unidimensionality

Unidimensionality is an important measurement used to evaluate if the developed instruments are able to measure what they are supposed to measure, in this case it is the instrument construct of student nationalism instruments. Rasch Model Analysis applies Principal Component Analysis from the residue to measure the variety of the instruments measuring what they should measure (Misbah & Sumintono, 2014).

The importance of building unidimensional structures is to provide evidence of internal consistency (Curtis & Boman, 2007; Huberty, Vener, Gao, Matthews, Ransdell, & Elavsky, 2013)

Table 7. Dimensionality Item of 420 Respondents

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units		Eigenvalue	Observed	Expected
Total raw variance in observations	=	64.1128	100.0%	100.0%
Raw variance explained by measures	=	22.1128	34.5%	35.7%
Raw variance explained by persons	=	6.1511	9.6%	9.9%
Raw variance explained by items	=	15.9617	24.9%	25.7%
Raw unexplained variance (total)	=	42.0000	65.5%	100.0%
Unexplned variance in 1st contrast	=	4.5508	7.1%	10.8%
Unexplned variance in 2nd contrast	=	2.0455	3.2%	4.9%

Table 7 above presents the data that the result of raw variance measurement is 34.5%. It means that it gets across the unidimensionality minimal requirement of 20%. Accordingly, the unexplained variance is the variance which is ideally no more than 10%. According to Sinnema et al. (2016) the criterion for establishing Unidimensionality is that the raw variance explained by the size must explain at least 20 percent of the test variance. This shows that Unidimensionality requirements have been met. Another thing, unexplained variance is the variance that cannot be explained by the instrument ideally does not exceed 10%. (results show all variance data below 10%). This indicates that the 42-point statement of student nationalism instruments can map students' perceptions about student nationalism, on condition that this can fulfill the constructed dimensions and indicators.

Conclusion

The analysis results of the developed instruments can be used as an empirical support to state that the instruments of student nationalism measurement has good psychometric guarantee. This can be seen through the result of Cronbach Alpha Reliability (KR-20) reaching the score of 0.89 with item reliability up to 0.99. Generally, respondents have high level of nationalism. Referring to the instruments of unidimensionality, the result shows that measurement is able to explain 34.5% of variance which appears in group of respondent.

Among 50 numbers of question/statement items, there are 35 final items after the process of Rasch Model Analysis Phase 1 (8 items eliminated) and 2 (7 items eliminated).

Reference

- Abdullah, N., & Lim, B. K. (2013). Parallel Circuit Conceptual Understanding Test (PCCUT). *Procedia-Social and Behavioural Sciences*, 90, 431-440. Doi: 10.1016/j.sbspro.2013.07.112
- Adisusilo, S. (2009). Nationalism-Democracy_Civil Society. <http://www.usd.ac.id/lembaga/lppm/.../vol.23.No.2> October 2009. (accessed in March 22, 2017)
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey: Brooks/Cole publishing company.
- Anshel, M. H., Weatherby, N. L., Kang, M., and Watson, T. (2009). Rasch Calibration of a Unidimensional Perfectionism Inventory for Sport. *Psychology of Sport and Exercise*, 10, 210-216.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model* (3rd ed.). New York: Routledge.
- Bybee, R., McCrae, B., & Laurie, R. PISA (2006). An Assessment of Scientific Literacy. *Journal of Research in Science Teaching*, 46(8), 865-883. 2009. <https://doi.org/10.1002/tea.20333>.
- Curtis, D. D., & Boman, P. (2007). Xray your data with Rasch. *International Education Journal*, 8(2), 249-259.
- Gable, Robert K. dan Marian B. (1993). *Wolf, Instrument Development In Affective Domain*. Boston: Kluwer Academic Publisher.
- Gómez, Laura E., Arias B., Verdugo M.A., and Navas P. (2012). Application of the Rasch Rating Scale Model to The Assessment of Quality of Life of Persons with Intellectual Disability. *Journal of Intellectual dan Developmental Disability*,; 37(2): 141–150
- Haiyang, S. (2010). An Application of Classical Test Theory and Many-facet Rasch Measurement in Analysing the Reliability of an English Test for Non-English Major Graduates. *Chinese Journal of Applied Linguistics (Bimonthly)*, 33(2). 87-102. 2010.
- Hambleton, R. K., & Jones, R. W. (1993). *Comparison of Classical Test Theory and Item Response Theory and Their Applications to the Test Development*.
- Huberty, J., Vener, J., Gao, Y., Matthews, J. L., Ransdell, L., & Elavsky, S. (2013). Developing an Instrument to Measure Physical Activity Related Self-Worth in Women: Rasch analysis of the Women's Physical Activity Self-Worth Inventory (WPASWI). *Psychology of Sport and Exercise*, 14(1), 111-121. Doi: 10.1016/j.psychsport.2012.07.009.
- Kartodirjo, S. Kartodirjo, S. (1999). Nationalism, Formerly and Nowadays, the Dynamics of Indonesia Nationalism. eds. Dance I. Palit et.al. Salatiga:Yayasan Bina Darma. eds. Dance I. Palit et.al. Salatiga:Yayasan Bina Darma.
- (1999). *Multi-dimentional of Nation-building, Nationalism Ethics, and Unitary State*. Yogyakarta: Kanisius.
- Kunaifi, A. & Ayu, F. P. (2012). Effectiveness of Civic Education Course in Developing Accountant Students' Nationalism Attitude in Dealing with IFRS Implementation. A paper presentation, presented in Indonesia National Conference of Accountant Education.
- Linacre, J. M. (2004). Rasch Model Estimation: Further Topics. *Journal of Applied Measurement*, 5.
- Linacre, J.M. (2009). *Winsteps (Version 3.68.0) [Computer Software and manual]*. Chicago: Winsteps.com.
- Misbah, I.H & Sumintono, B. (2014). Instrument Development and Validation "Students' Perception on Teachers Morality" in Indonesia using Rasch Model, presented in National Seminar on "Development of Valid Character Assessment Instrument" at Faculty of Psychology of Universitas Muhammadiyah Surakarta.
- Naga, D. S. (2013). *Theory of Scoring on Mental Measurement*. Jakarta: Nagrani Citrayasa.
- Perera, C.J., Sumintono, B., & Na, Jiang. (2018). The Psychometric Validation of The Principal Practices Questionnaire Based on item response theory. *International Online Journal of Educational Leadership*. 2(1), 21-38.
- Retnawati, H. (2014). *Theory of Response Item and Its Implication*. (1st ed.). Yogyakarta: Parama Publishing.

- Shafer, B. C. (1955). *Nationalism, Myth and Reality*. New York: A Harvest Book Harcourt.
- Smith, A. D. (2003). *Nationalism Theory, Ideology, History*. Jakarta:Erlangga.
- Sinnema, C., Ludlow, L., & Robinson, V. (2016). Educational leadership effectiveness: a Rasch analysis. *Journal of Educational Administration*, 54(3), 305 - 339.
- Sofyan, Silvia, F., & Sundawa, D. (2015). "Correlation between Civic Education Course and Students' Improvement in Nationalism Spirit and National Understanding". *JPIS, Journal of Social Science*, Vol 24, No. 2
- Sumintono, Bambang, & Widhiarso, W. (2014). *Application of Rasch Model in Scientific Research*, revised edition. Cimahi:Trim Komunikata Publishing House.
- Sumintono, B. (2017). "The Rasch Model for Educational Assessment: An Introduction." National Teachers and Lectures Conference (KGDN). Medan: Auditorium Sumatera Utara (USU).
- Wafa, H. & Satmoko, A. (2017). "An Effort of the Indonesian Moslem Student Movement (PMII) of Surabaya State University Commissariat to Nationalism Building". *Moral and Citizenship Studies*, Volume 05 No. 03 Jilid III.
- Wibisono, S. (2015). Applying the Rasch Model in Validation of Religion Fundamentalism Measurement Instrument for Moslem Respondents," *Journal of Indonesian Psychology and Education Measurement*, Vol IV.