
Norming of Coloured Progressive Matrices Test in Elementary School Children Based on Classical Measurement Theory and Rasch Modeling

Uly Gusniarti, Mira Aliza Rachmawati, Susilo Wibisono,
Libbie Annatagia, Ike Agustina, Rumiani

Department of Psychology, Universitas Islam Indonesia, Indonesia

uly.gusniarti@uii.ac.id

Abstract

This study aimed to develop Coloured Progressive Matrices (CPM) norms for the use in the Indonesian context. We used two approaches, namely classical test theory (CTT) which uses raw score (total score) as measurement result information and Rasch modeling which uses logit value as measurement result information. This research was conducted in four regencies and one municipality in the Province of Yogyakarta. The participants were 1,779 elementary school age children recruited through random sampling. The norming analysis in this study divided the data into five age groups in the range of 6 – 12.5 years old. The level of intelligence represented by the results of the CPM measurement consists of five levels, from Grade I to Grade V. Grade V as the lowest intelligence level has a value below the 5th Percentile of the data distribution. Grade IV as the second lowest level of intelligence was located between between the 5th and 25th Percentile of the data distribution. Grade III representing the average level of intelligence had the greatest range from 25th to 75th Percentile. In addition, the range allocated for Grade II was similar to Grade IV, but in the opposite direction of the distribution (i.e., between 75th and 95th Percentile). Lastly, Grade I as a representation of the highest level of intelligence is in the range of values above the 95th Percentile.

Keywords: CPM, intelligence, norming, elementary school, Indonesian context.

Abstrak

Penelitian ini bertujuan untuk mengembangkan penormaan terhadap hasil tes inteligensi CPM yang relevan dalam konteks Indonesia. Basis pendekatan yang digunakan dalam penelitian ini adalah teori pengukuran klasik yang menggunakan skor mentah (skor total) sebagai informasi hasil pengukuran dan pemodelan Rasch yang menggunakan nilai logit sebagai informasi hasil pengukuran. Penelitian ini dilakukan di lima daerah di D. I Yogyakarta dengan jumlah subjek sebanyak N=1779 siswa Sekolah Dasar. Analisis penormaan yang dilakukan membagi data ke dalam lima kelompok usia yang bergerak antara 6 – 12,5 tahun. Level kecerdasan yang direpresentasikan oleh hasil pengukuran CPM dibagi ke dalam lima level yaitu Grade I – V. Grade V sebagai level kecerdasan paling rendah berada dalam rentang nilai di bawah Persentil ke 5. Grade IV di atasnya berada antara nilai Persentil ke 5 – Persentil ke 25. Grade III di atas Grade IV berada antara nilai Persentil ke 25 – Persentil ke 75. Grade II di atas Grade III berada di antara rentang nilai Persentil ke 75-Persentil ke 95. Grade I sebagai representasi tingkat inteligensi paling tinggi berada pada rentang nilai di atas Persentil ke 95.

Kata Kunci: CPM, inteligensi, penormaan, sekolah dasar, konteks Indonesia.

Introduction

The measurement of intelligence domain, both in general and in specific terms, has an important value. This is mostly associated with the efforts to map individual cognitive potential which is often used as a reference in various aspects of the learning process. Currently, there are many tests to measure individual intelligence scientifically developed and practically used by psychologists in many fields. One of the tests widely used in Indonesia is Raven's Progressive Matrices (RPM). This test represents a nonverbal approach to measure individual cognitive ability. It has been used largely in Indonesia for many reasons, mainly because the non-verbal nature of the test can avoid the potential cultural biases. RPM consists of three versions, each of which has a different focus of use. They are: (1) Standardized Progressive Matrices (SPM) which the original form was firstly published in 1938. The test consists of five sets (A-E), each containing 12 items representing different levels of difficulty, from easy to difficult; (2) Colored Progressive Matrices (CPM) which was designed for children aged 5-11 years, the elderly, and individuals with physical or mental disabilities. This test consists of a set A and a set B of the SPM, and a set of inserts known as the Ab set. Some items are presented in color to visually stimulate the test-takers; and (3) Advanced Progressive Matrices (APM) which was designed for teenagers and adults with above average level of intelligence. This test consists of two sets with 48 items for set I and 36 items for set II.

The RPM test consists of pictorial questions, in the form of a large picture with holes and under the picture there are six or eight small pictures as choices to answer and complete the question. Individuals tested with this tool were asked to choose the most appropriate choice to close the hole in the big picture (Azwar, 2002). As a complete component of a test tool, the norms of a test tool as a guide in interpreting test results are important. Norms in intelligence tests are tools to interpret the results obtained from a measurement/test (Domino & Domino, 2006). The norming of test tool that can be used as a broad reference is a necessity in psychodiagnosics research. In the Indonesian context, this is one of the important tasks, considering that in the practice of psychological measurement in Indonesia, the norms that are usually used are norms developed with reference to subjects from outside Indonesia. Referring to the existing discrepancy, these norms are often irrelevant and inappropriate to the conditions of people in Indonesia. The standardization of scores obtained from measurement through a norm is a mechanism to facilitate interpretation of the measurement results (Aaron, Coups, & Aaron, 2013). The sample in the norm must represent the type of individual being tested. If a measurement is made, for example, for the purpose of evaluating student performance in a certain environment and does not use local norms as a reference, it is possible that an evaluation and decision-making process will occur that is not in accordance with the geographical or institutional conditions of the respondent (Urbina, 2004).

RPM (Raven Progressive Matrices) was developed by Raven with reference to Spearman's theory of intelligence which states that there are two main components in general cognitive abilities (g factor), namely eductive and reproductive abilities (Raven, 2000). Eductive ability refers to the ability to produce a meaningful picture of something confusing and the ability to solve a high level of complexity in nonverbal aspects (Raven, 2000). Reproductive ability refers to an individual's ability to absorb, remember and reproduce explicit information and communicate it to another person (Raven, 2000). Basically, the RPM test consists of pattern questions that have missing pieces. Subjects who have taken the test are asked to choose pattern pieces that can complete the existing pattern (Raven, 2000).

RPM (Raven Progressive Matrix) has undergone several revisions, which in the end emerged APM (Advanced Progressive Matrix) and CPM (Colored Progressive Matrix). Both aim to cover the shortcomings of the initial series of RPM (Raven Progressive Matrices) tests, namely Raven's Standard Progressive Matrices (SPM), related to groups of subjects who have low abilities and subjects who have high abilities. APM (Advanced Progressive Matrices) aims to facilitate groups of subjects who have high abilities, for example highly educated adults. Meanwhile, CPM (Colored Progressive Matrices) can

facilitate groups of subjects who have low abilities as well as the group of children (Raven, 2000). CPM is used in populations of children, late adults and people with disabilities (Smits, Smit, Heuvel, & Jonker, 1997). The CPM consists of Sets A, Ab, and B, each of which has 12 items and is designed for children aged between 5–11 years. The form of the progressive matrix test (RPM) has previously been used to measure intelligence in children.

The previous form of the progressive matrices test (RPM) has been used to measure intelligence in children. However, the response that emerged from these children was different from the response in adults. Adults are able to understand what is expected of the test even without listening to the instructions, but children have difficulty with this. Based on this, Raven then compiled another version of the progressive matrices he developed, namely CPM (Raven, 2000). The main difference between CPM and other versions is that CPM is presented in colors, not just black and white. The interpretation of intelligence test results using CPM is explained through grade criteria or levels, which consist of 5 levels, namely:

- Grade I: Intellectually superior
- Grade II: Intellectual capacity above average
- Grade III: Average/Normal
- Grade IV: Intellectual capacity below average
- Grade V: intellectual retardation

The development of this classification or norm is carried out using percentiles (P) (Raven, 2000), namely P95-P99 for Grade I, P75 – P95 for Grade II, P25 – P75 for Grade III, P5 – P25 for Grade IV, and P1- P5 for Grade V. Various norms have been developed to find interpretations that relevant to a particular population. Raven (2000) himself has carried out a norm with a sample of 291 children aged 5-10 from schools in Dumfries. In the second stage, Raven (2000) also developed norms involving 608 children aged 5 – 11 years from schools in Dumfries, Scotland. In principle, norming is an attempt to facilitate the interpretation of test results. Azwar (2010) revealed that the measurement results in the form of numbers require a comparison norm in order to be interpreted qualitatively. Basically, the interpretation of psychological test scores is always normative. This means that the score refers to the relative position of the score on a pre-defined group. This can be done, among others, with the help of descriptive statistics from the distribution of group score data which generally includes the number of subjects, average, maximum score and minimum score (Azwar, 2010).

The use of classical theory of measurement and Rasch modeling refer to the theoretical differences between the two. The fundamental difference between the Rasch model and the classical measurement theory lies in how to treat raw scores in the analysis process. In classical measurement theory, the raw score in the form of a rating scale is directly analyzed and treated as data as if it had an integer character. Whereas in the Rasch Model, raw data cannot be directly analyzed, but must first be converted into the form of 'odds ratio'. Then the logarithmic transformation is carried out into logit units as a manifestation of the respondent's probability when responding to an item. Referring to this procedure, Sumintono and Widhiarso (2013) stated that the Rasch model can be used as a method of returning data according to its natural condition. This natural condition refers to the basic characteristics of quantitative data, which is continuum. A classical measurement theory that uses raw data from the response of a rating is considered unable to present the original characteristics of quantitative data that is a continuum. Through the Rasch model, an ordinal response can be transformed into a ratio that has a higher level of accuracy, with reference to the probability principle.

Methods

This research uses a survey method which is conducted extensively. This study aims to compile the norm of a measurement tool that is adapted from different cultural contexts, thus requiring a large number of research subjects. Therefore, the most relevant method in this context is the survey method. Survey is a data collection mechanism that is carried out on a sample group that represents a certain population. The sample involved in this study was taken based on the number of elementary schools in the Province of the Special Region of Yogyakarta. Sample representation was carried out through a sampling equally in four regencies and one municipality in the Special Region of Yogyakarta. Elementary school data were taken from the website of the Education Office of the Special Region of Yogyakarta. Elementary school criteria are determined by location and level of accreditation. Each regency/municipality is represented by one primary school with an A accreditation, one primary school with B accreditation, and one primary school with C accreditation.

This study uses a randomization technique with a table of random numbers. Based on this process, the sampling conducted in this study can be regarded as a stratified cluster random sampling. The location of this research was obtained based on a sampling process, namely in Sleman, Kulonprogo, Bantul, Gunungkidul, and Yogyakarta. Elementary school location data obtained based on the sampling process are as follows:

Table 1. Elementary School Sample

Location	Name of School	Address	Accreditation
Sleman	SD Negeri Kaweden	Mlati, Sleman	A
	SD Negeri Sendangharjo	Padon, Minggir, Sleman	B
	MI Maarif Gerjen	Gerjen, Seyegan, Sleman	C
	SD Negeri Kranggan	Jl. Daendels, Galur, Kulon Progo	A
Kulonprogo	SD Negeri 3 Brosot	Pulo, Galur, Kulon Progo	B
	SD Negeri Selo	Kokap, Kulon Progo	C
	SD Jaranan	Pringgolayan, Banguntapan, Bantul	A
Bantul	SD Rejodadi	Brengosan, Kasihan, Bantul	B
	SD Kembangan	Kembangan. Bambanglipuro, Bantul	C
	SD Negeri Gupit	Ngipik, Gedangsari, Gunung Kidul	A
Gunung Kidul	SD Negeri Hargomulyo 1	Gedangsari, Gunung Kidul	B
	SD Negeri Tancep II	Sendangrejo, Ngawen, Gunung Kidul	C
	SD Negeri Serayu	Gondokusuman, Yogyakarta	A
Kota Yogyakarta	SD Negeri Ngabean	Ngampilan, Yogyakarta	B
	SD Negeri Tukangan	Pakualamanan, Yogyakarta	C

Evaluation of the validity of the measurement uses two reviews, namely the response process and external criteria. Evaluation of validity based on the response process is carried out by conducting tests based on the guidelines of the existing test tool and carried out under the supervision of a psychologist. Meanwhile, the evaluation of the validity based on external criteria is carried out using the age of the subject and school accreditation as criteria. The theoretical assumption built in this validation process is that subjects from schools that have A and B accreditation will have higher scores than subjects from schools

with C accreditation. The older the subject, the higher the score obtained. Based on age criteria, theoretical assumptions are used to test the validity that the older the subject, the higher the score obtained.

The score used in this validation process is the logit person value obtained from the process of calculating the logit value in the Rasch model. Meanwhile, comparisons between groups, both based on accreditation (A, B, and C) and age criteria, were carried out using one-way analysis of variance (ANOVA).

The data analysis technique used in this research is descriptive analysis to find various criteria needed in the norming of measuring instruments. Some of the techniques used include cross-tabulation based on age, scores for each component, and number of children. Demographic data such as family background, socioeconomic level, school achievement scores will also be used in the data analysis. Norming was carried out using percentile values, as was the procedure previously developed by Raven (2000). After carrying out these procedures, the CPM norm is obtained, which can be used in interpreting the CPM score for children aged 6-13.

Description of Research Subject

The number of students who became the subject of this study were 1779 elementary school students in the province of the Special Region of Yogyakarta. As planned for this research, the subjects come from four regencies and one municipality in the Special Region of Yogyakarta and from schools that have A, B and C accreditations from the Education Office of the Special Region of Yogyakarta. Based on gender, the research subjects consisted of 912 male students (51.26%) and 867 female students (48.74%). Based on their age, the distribution of research subjects is presented in Table 2. below:

Table 2. Distribution of Research Subjects Based on Age

Age	Frequency	Percentage
6 – 7.5 years old	321	18.04%
7.5 – 8.5 years old	349	19.62%
8.5 – 9.5 years old	333	18.72%
9.5 – 10.5 years old	406	22.82%
10.5 – 11.5 years old	276	15.51%
11.5 – 12.5 years old	94	5.28%
	1779	100.00%

The age distribution of the research subjects tends to be equally distributed, which is between 6 to 12.5 years. Based on the location and the level of school accreditation, the data obtained are presented in Table 3 below:

Table 3. Distribution of Research Subjects Based on Location and Level of School Accreditation

		Accreditation			Total
		A Accreditation	B Accreditation	C Accreditation	
Regency/ Municipality	Sleman	114	78	15	207
	Yogyakarta	330	178	275	783
	Bantul	134	132	89	355
	Kulon Progo	118	53	35	206
	Gunung Kidul	72	55	101	228
Total		768	496	515	1779

From table 3 above, it is known that the distribution of research subjects is quite wide. The distribution of subjects is also representative to describe the psychological characteristics of elementary school age children in Indonesia, especially in the Special Region of Yogyakarta Province.

Results and Discussion

Instrument Validity Evaluation

Validity evaluation is needed as an assurance that the measurement instrument measures the general intelligence construct in children appropriately. Validity is not only related to the instrument, but also related to the data collection process. Validity is a crucial concept in quantitative research because it provides an assurance of conformity between the theoretical concepts that are the basis of the empirical evidence represented by the data (Purwono, 2014). This study uses two main validity assurances, namely the validity based on the response process and the validity based on external criteria. Response-based validity provides assurance that the instrument filling process carried out by research subjects is in accordance with what was intended by the researcher. This is guaranteed by carrying out standardized procedures in the filling instructions. In addition, data collection is carried out by professional staff, Master of Professional Psychology students, who have carried out the Professional Psychologist Work Practice. Testers give standard instructions in accordance with the test manual. Tester provides and explains examples and how to answer it. Respondents were also given the opportunity to practice and the tester makes sure the answer is correct. The tester also gives the opportunity for the respondent to ask questions if they still don't understand the instructions on how to take the test. Therefore, based on the response process, the research data obtained in this study are valid.

The second review, evaluation of validity based on external criteria, was carried out based on the school accreditation variable and the age variable of the research subject. This validity evaluation uses a score in the form of the respondent's logit value obtained through the analysis of the Rasch model.

The hypotheses built to justify the validity based on these two criteria are:

- There is a difference in the CPM score (logit value) of the research subjects based on the school's accreditation level. Subjects from schools with A and B accreditation levels will have a higher CPM score than subjects from schools with C accreditation levels.
- There is a difference in the CPM score (logit value) of the research subjects based on their age level. Research subjects with a higher age level will have a higher CPM score than subjects with a lower age level.

This validity evaluation was carried out in two stages, namely: (1) Calculating the score of each research subject using the logit person value in the Rasch model; and (2) Conducting an analysis of differences between group based on school accreditation and age using 1-way analysis of variance (ANOVA).

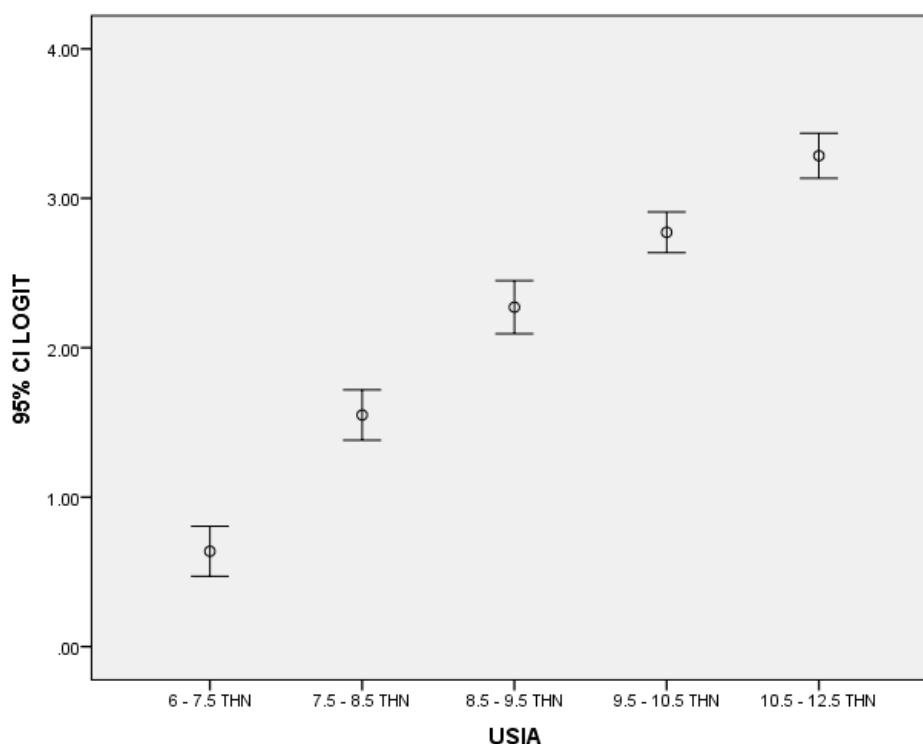
The results of ANOVA to test the first hypothesis are shown in Table 4 below:

Table 4. Difference in Score based on School Accreditation Level

Analysis of Differences	F	p	Description	LSD Analysis	Mean Difference	p	Result
Subject Score based on School Accreditation Level	16.371	0.00 (p < 0.01)	A Significant Difference	A-B	0.102	0.31 (p>0.05)	No difference
				A-C	0.547	0.000 (p<0.01)	significant
				B-C	0.446	0.000 (p<0.01)	significant

Based on Table 4. above, it is known that the F value from the subject logit value = 16.371 with $p = 0.000$ ($p < 0.01$). This shows that there is a significant difference in scores on research subjects based on the level of school accreditation. Subjects from schools that have A and B accreditations are shown to have higher scores than subjects from schools with C accreditation level. The difference in scores between the subject groups of schools that have A accreditation and schools that have C accreditation is 0.547 ($p = 0.000$). The difference in subject scores from schools that have B accreditation and subjects from schools that have C accreditation is 0.446 ($p = 0.000$). This shows that students from schools with A and B accreditation levels have higher scores than students from schools with C accreditation level. This information can be used as a validity argument based on the first external criteria.

The second hypothesis, which serves as an assurance of the validity of the instrument, was analyzed using ANOVA. Visually, the results of the analysis are shown in Figure 3. below:



Source: Personal Data.

Figure 1. Analysis of Differences Based on Age Group

Based on the results of one-way ANOVA, the value of $F = 163,476$ with $p = 0.000$ ($p < 0.01$) was obtained. This shows that there is a significant difference in scores based on the age group that is the subject of the study. Figure 3. above shows that the higher the age level of the research subject, the higher the score obtained. This can be used as an argument to prove that the instrument and the data collection process are valid.

One of Rasch's assumptions is unidimensional and local independence. The unidimensional assumption is that the ability measured by the items is a single thing. Ideally, each item measures only one ability or psychological aspect of the respondent, not measuring two or more respondents' attributes. The analysis of the Rasch model using Principal Component Analysis of the residuals shows that the raw data variance measurement results are 49.3%. According to Sumintono and Widhiarso (2013), the minimum unidimensional requirement is 20%. In this study, the value is more than 40%. It means that

it is better than the minimum standard. In addition, the variance that cannot be explained by the instrument should ideally not exceed 15% and in this study the unexplained variance 1st contrast = 11.4%, unexplained variance 2nd contrast = 3.5%, unexplained variance 3rd contrast = 2.0%, unexplained variance 4th contrast = 1.9%, and unexplained variance 5th contrast = 1.8% (Table 5)

Table 5. Standardized Residual Variance (in Eigenvalue units)

Description	Empirical (%)	Model
variance explained by measures	49.3 %	67.5 %
unexplained variance (total)	50.7 %	32.5 %
unexplained variance in 1st contrast	11.4 %	
unexplained variance in 2nd contrast	3.5 %	
unexplained variance in 3rd contrast	2.0 %	
unexplained variance in 4th contrast	1.9 %	
unexplained variance in 5th contrast	1.8 %	

Meanwhile, the assumption of local independence means that the subject's response to the item has no effect on the response to other items. The assumption of local independence will be fulfilled if the respondent's answer does not depend on the answer to other items. Based on the correlation of the largest standardized residuals, there are pairs of items that have a correlation of more than 0.7. However, all items in this instrument are still analyzed according to their original instrument.

Table 6. Largest Standardized Residual Correlations

Correlation	Item	Item
0.86	1 A1	2 A2
0.85	1 A1	3 A3
0.81	1 A1	25 B1
0.78	1 A1	4 A4
0.76	1 A1	15 AB3
0.73	2 A2	3 A3
0.72	3 A3	25 B1
0.70	3 A3	4 A4

Analysis of Norming

In this study, norming process is carried out by referring to the norming model developed by Raven (2000), by using percentile values to divide the subject into several values. Grades P95-P99 are used as a standard for Grade I (intellectually superior), P75 – P95 for Grade II (above average intellectual capacity), P25 – P75 for Grade III (Average or Normal), P5 – P25 for Grade IV (Intellectual capacity below average), and P1-P5 for Grade V (Intellectual retardation). Norming process is carried out in each age category of research subjects. Furthermore, norming process is carried out using two types of values, the logit value of the subject generated from the Rasch model and the total value generated from the sum of the item scores answered correctly by the subject. The first categorization is made into 14 age categories among the ages of 6.5 years to 13 years. The results of norming process based on the age range are shown in Table 7. below:

Table 7. Norming Results in 14 Age Categories

Age	N	Logit Value					Total				
		P ₅	P ₂₅	P ₅₀	P ₇₅	P ₉₅	P ₅	P ₂₅	P ₅₀	P ₇₅	P ₉₅
6.5	40	-2.49	-0.81	0.14	1.01	2.41	8.15	14.25	18.50	22.75	28.85
7	130	-1.64	-0.63	0.24	1.1125	2.58	11.00	15.00	19.00	23.25	29.45
7.5	148	-1.11	0.03	1.06	1.94	3.69	13.00	18.00	23.00	27.00	32.55
8	190	-1.37	0.45	1.71	2.45	3.89	12.00	20.00	26.00	29.00	33.45
8.5	155	-0.92	0.65	1.49	2.74	4.50	13.80	21.00	25.00	30.00	34.00
9	151	-0.40	1.27	2.45	3.44	5.46	16.00	24.00	29.00	32.00	35.00
9.5	160	-0.18	1.27	2.19	3.06	4.50	17.00	24.00	28.50	32.00	34.00
10	213	0.03	1.71	2.74	3.44	4.50	18.00	26.00	30.00	33.00	34.30
10.5	179	0.65	1.94	3.06	3.89	5.46	24.00	28.00	31.00	34.00	35.00
11	177	1.27	2.19	3.44	4.50	5.46	24.00	28.00	32.00	34.00	35.00
11.5	99	0.65	2.74	3.44	4.50	5.46	21.00	30.00	32.00	34.00	35.00
12	49	0.34	2.07	3.06	4.50	5.46	19.50	27.50	31.00	34.00	35.00
12.5	18	0.45	1.83	3.25	4.74		20.00	26.50	31.50	34.25	
13	22	1.16	2.19	3.06	3.55	4.50	23.45	28.00	31.00	32.25	34.00

Table 7. above shows the values used for norming process based on 14 age categories among 6.5 years to 13 years. However, in the norming model above, there are still things that deviate from the assumptions.

This is indicated by the presence of smaller percentile values at higher ages, such as P5 values at 11, 11.5 and 12 years old or P25 values at 11.5 years, 12 years and 12.5 years. Based on the problem of consistency, it is necessary to formulate a second norming model that uses five age categories.

The norming results based on the five age categories are presented in Table 8. below:

Table 8. Norming Results in Five Age Categories

Age	N	Logit Value					Total				
		P ₅	P ₂₅	P ₅₀	P ₇₅	P ₉₅	P ₅	P ₂₅	P ₅₀	P ₇₅	P ₉₅
6 - 7.5	318	-1.64	-0.40	0.65	1.49	3.08	11.00	16	21.00	25.00	31.05
7.5 - 8.5	345	-1.11	0.65	1.71	2.45	4.50	13.00	21	26.00	29.00	34.00
8.5 - 9.5	310	-0.28	1.27	2.19	3.44	4.50	16.55	24	28.50	32.00	34.00
9.5 - 10.5	392	0.45	1.94	2.74	3.89	4.50	20.00	27	30.00	33.00	34.35
10.5 - 12.5	366	0.79	2.19	3.44	4.50	5.46	21.70	28.00	32.00	34.00	35.00

The norming results in Table 8. above are consistent with the assumption, that in childhood, the higher the level of the age, the more the average logit value or the average IQ score of the child increases. Based on the results of the analysis, a norm is formulated based on five criteria of children age. The norming results are presented in Table 9. below:

Table 9. Norming Results for Each Grade

Grade	Age (years old)	Criteria	Norms Based on Total Score	Norms Based on Logit Value
Grade I	6.0 – 7.5	P ₉₅ - P ₉₉	$X \geq 31.05$	$\text{Log} \geq 3.08$
	7.5 – 8.5		$X \geq 34.00$	$\text{Log} \geq 4.50$
	8.5 – 9.5		$X \geq 34.00$	$\text{Log} \geq 4.50$
	9.5 – 10.5		$X \geq 34.35$	$\text{Log} \geq 4.50$
	10.5 – 12.5		$X \geq 35.00$	$\text{Log} \geq 5.46$
Grade II	6.0 – 7.5	P ₇₅ - P ₉₅	$31.05 > X \geq 25.00$	$3.08 > \text{Log} \geq 1.49$
	7.5 – 8.5		$34.00 > X \geq 29.00$	$4.50 > \text{Log} \geq 2.45$
	8.5 – 9.5		$34.00 > X \geq 32.00$	$4.50 > \text{Log} \geq 3.44$
	9.5 – 10.5		$34.35 > X \geq 33.00$	$4.50 > \text{Log} \geq 3.89$
	10.5 – 12.5		$35.00 > X \geq 34.00$	$5.46 > \text{Log} \geq 4.50$
Grade III	6.0 – 7.5	P ₂₅ - P ₇₅	$25.00 > X \geq 16.00$	$1.49 > \text{Log} \geq -0.40$
	7.5 – 8.5		$29.00 > X \geq 21.00$	$2.45 > \text{Log} \geq 0.65$
	8.5 – 9.5		$32.00 > X \geq 24.00$	$3.44 > \text{Log} \geq 1.27$
	9.5 – 10.5		$33.00 > X \geq 27.00$	$3.89 > \text{Log} \geq 1.94$
	10.5 – 12.5		$34.00 > X \geq 28.00$	$4.50 > \text{Log} \geq 2.19$
Grade IV	6.0 – 7.5	P ₅ - P ₂₅	$16.00 > X \geq 11.00$	$-0.40 > \text{Log} \geq -1.64$
	7.5 – 8.5		$21.00 > X \geq 13.00$	$0.65 > \text{Log} \geq -1.11$
	8.5 – 9.5		$24.00 > X \geq 16.55$	$1.27 > \text{Log} \geq -0.28$
	9.5 – 10.5		$27.00 > X \geq 20.00$	$1.94 > \text{Log} \geq 0.45$
	10.5 – 12.5		$28.00 > X \geq 21.70$	$2.19 > \text{Log} \geq 0.79$
Grade V	6.0 – 7.5	P ₁ - P ₅	$X < 11.00$	$\text{Log} < -1.64$
	7.5 – 8.5		$X < 13.00$	$\text{Log} < -1.11$
	8.5 – 9.5		$X < 16.55$	$\text{Log} < -0.28$
	9.5 – 10.5		$X < 20.00$	$\text{Log} < 0.45$
	10.5 – 12.5		$X < 21.70$	$\text{Log} < 0.79$

The norming process based on logit values is quite difficult to use pragmatically, because in its application the CPM test is used individually. So that the evaluation using the Rasch model based on the logit person value can be done by entering the test results into the sample data.

Discussion

This study aims to develop CPM test norms for children aged 6-13 years. 1779 elementary school children in the province of the Special Region of Yogyakarta were involved in this study. Norming analysis is carried out using two approaches as a comparison, namely an approach based on classical test theory and an approach based on Rasch modeling. The second approach is not based on the raw score generated in the measurement process, but based on the logit value obtained from each research subject.

The results of the evaluation of the validity based on external criteria indicate that this instrument is valid. The advantage of this instrument compared to other verbal intelligence tests is that CPM can eliminate the possibility of cultural and language bias in intelligence tests (Kazem, et al 2009). It is because the CPM test uses images so that it can be understood equally by individuals from different cultures and languages.

The norming criteria in this test use five grades or levels, grade I to grade V. Grade I indicates a very high level of intelligence for children. Grade II indicates a high level of intelligence, and grade III indicates an average intelligence level. Justification for intelligence capacity in children aged 6-13 years can be seen from the total score of the CPM test and is included in the criteria for each grade based on age.

This is because the principle of measurement results that can be labeled on individuals cannot be done by adding up the item scores (Sumintono & Widhiarso, 2013). This addition has a fundamental limitation because the scores that are added up basically do not meet the basic criteria for integers, so the scores cannot be subjected to arithmetic operations. The logit value (logarithmic odds unit) is basically a representation of the individual's probability of answering the test items. Therefore, in individual practical use, the logit value cannot be obtained because there is no comparison group.

Then how is the use of norms developed using Rasch modeling in the context of individual practical measurements? The researcher suggests that data from 1779 of these subjects be used as a benchmark for determining the logit value of the new respondents. The raw score on each item generated by the respondent is entered into a list and then the logit value is seen based on the 1779 subjects. From the results of the logit value, the tester can determine what grade the respondent is in, based on his chronological age.

Previously, Raven (2000) had conducted several standardization studies on the CPM test, including the standardization conducted in 1992 in Dumfries, Scotland and the 1993 standardization conducted in Des Moines, Iowa. This study is still possible to be improved by adding research subjects in the age group with a smaller sample size. As presented in Table 5, there are several age categories where the number of subjects is still relatively small. The age categories include the age group under 7 years old and the age group above 11 years old. However, by using a wider range of categorization as shown in Table 6, the limitations related to the number of subjects can still be overcome.

Conclusion

Based on the results of data analysis on 1779 research subjects aged 6-13 years, a standardization for the CPM intelligence test instrument was obtained using five age categories and five grades. The five age categories are ages 6 – 7.5 years, 7.5 years – 8.5 years, 8.5 years – 9.5 years, 9.5 years – 10.5 years and lastly, 10.5 years – 12.5 years. Meanwhile, the grading uses the CPM pattern, grade I for the highest intelligence level, and followed by levels below it up to grade V. The norming analysis is presented in Table 4.8, which contains the norming based on the raw score and the norming based on the logit value generated from the Rasch Model.

Researchers suggest that the results of this norming are used in the interpretation of measurements of subjects who come from regions in Indonesia that have relatively similar socio-demographic conditions with the province of the Special Region of Yogyakarta. This refers to the wide socio-demographic variance that exists in various regions in Indonesia. However, the CPM test is claimed to be a test that is free from the possibility of cultural and linguistic bias because it uses a tool in the form of images. This tool is seen as something that is more universal than numerical and verbal tools. In the practical use of norms that are carried out in the logit value, the researcher suggests that the data of 1779 subjects of this study become the database to generate logit scores. The simplest way in this process is to enter the respondents' data into the subject's data line and perform an analysis to find the respondent's logit value. Furthermore, the interpretation of the logit value can be carried out using the norms that developed in this study.

References

- Aaron, E; Coups, J.E, Aron, E.N. (2013). *Statistic for psychology*, 6th edition. New Jersey: Pearson Education Inc.
- Azwar, S. (1996). *Penyusunan skala psikologi*. Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2002). *Tes inteligensi: sebuah pengantar*. Yogyakarta: Pustaka Pelajar
- Azwar, S. (2010). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar
- Domino, G & Domino, M.L. (2006). *Psychological testing: an introduction*. Cambridge: Cambridge University Press.
- Kazem, A.M. dkk. (2009). A normative study of the coloured progressive matrices test for Omani Children age 5 – 11 years. *Jurnal Pendidikan Malaysia*. 34 (1).
- Purwono. U. (2014). Validitas pengukuran. Yogyakarta: Materi Workshop Pengukuran Psikologi, diadakan oleh Prodi Psikologi UII.
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41, page: 1-48.
- Smits, C.H.M; Smit, J.H; Heuvel, N & Jonker, C. (1997). Norms for abbreviated raven's coloured progressive matrices in an older sample. *Journal of Clinical Psychology*, Vol: 53 (7), page: 687-697.
- Sumintono, B & Widhiarso, W. (2013). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. Jakarta: Tim Komunikata Publishing House.
- Urbina, S. (2004). *Essential of psychological testing*. New Jersey: John Wiley & Sons, Inc.