

Developing an instrument of national examination of equivalency education Package C of mathematics subject

*¹Ian Harum Prasasti; ²Edi Istiyono

¹Sekolah Tinggi Manajemen Informatika dan Komputer Kalirejo (STMIK Kalirejo)
Jl. Jend. Sudirman, Kalirejo, Kabupaten Lampung Tengah, Lampung 34174, Indonesia

²Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

*Corresponding Author. E-mail: ihp.harum8@gmail.com

Submitted: 30 August 2017 | Revised: 26 April 2018 | Accepted: 17 May 2018

Abstract

The national examination of equivalency education is a competency test to equalize non-formal education with formal education. Departing from the importance of the quality and expectations of the national examination of equivalency education package C mathematics subjects and because the results are inseparable from the implementation process, developing an evaluation instrument to assess the implementation of the national examination of equivalency education package C mathematics subjects is important. The purpose of this study is to develop a suitable instrument for conducting an evaluation of the national examination implementation of the equivalency education package C mathematics subject. The respondents in this research are package C test takers in Bantul Regency, Yogyakarta. The data were analyzed by using SPSS 20.0 and Lisrel 8.54. The result of the analysis shows (1) based on the data obtained from respondents of try-out, the developed instrument is valid, reliable and qualified as a fit model; (2) components in the instrument of test takers is learning time, socialization, test materials, and test venue; and (3) the instrument have the validity value of > 0.40 and reliability coefficient of > 0.70 .

Keywords: *instrument, evaluation, national examination, equality education*

Introduction

In its effort to build and monitor the quality of education and to meet the needs of equity in the education aspect, the Indonesian government has continuously made a policy to develop the national standard competency test instruments. One of the efforts undertaken is through the provision of Government Regulation No. 19 of 2005, on National Education Standard in Article 3 that is a basis for the planning, implementation, and supervision of education in order to realize a quality national education.

The nationally standardized test of competence, or commonly known as national examination, aims to conduct coaching and provide assistance to schools in an effort to improve the quality of education (Mardapi &

Kartowagiran, 2009). In addition, the toughest goal of the national examination is to improve clarity, efficiency, and also effectiveness in making decisions (Adow, Alio, & Thinguri, 2015). National examination also aims to measure students' learning achievement in certain subjects that are grouped into science and technology to assess the achievement of the national education standards (Mudjijanti, 2011).

The competency test is in the form of national examination and national examination of equivalency education. The use of the term national examination of equality education is due to the position of the exam results which can be accounted for an equivalent to the results of formal education exams. The purpose of the national examination is for the mapping of the quality of schools, the selec-

tion of entry into the next level of education, and provision of schools in an effort to improve the quality of education. It can also be categorized as a diagnostic test (Setiadi, et al. 2011). The national examination of equivalency education is a competency test to equalize non-formal education in the form of Package A equivalent to elementary school, Package B equivalent to junior high school, and Package C equivalent to senior high schools.

Package C as one of the national equivalency education examination programs is aimed to solve educational problems that cannot be coped by formal education. Some factors in the non-formal education which have not been solved include a problem in senior high schools, traumatic experience, school drop-outs, and hyperactive and autism children. Thus, for equivalency of the non-formal education with formal education, the government runs programs of the national equivalency examination.

The term 'national equivalency examination' is used since the result of equivalency examination is credible and accountable, and its position is equivalent to the result of national examination of formal education. Likewise, one of the efforts undertaken by the government through the provision of Law No. 20 of 2003 of Republic of Indonesia on National Education System in article 26 verse 6 explains that the result of non-formal education can be equivalent with the result of formal education program after going through an equivalent assessment process by institutes selected by the government. Then, the national examination for equivalency education participants will automatically get a certificate from a non-formal educational institutions such as the learning group of Package C (Raharjo, 2012).

Every educational activity needs an evaluation activity to know the level of success of the implementation of the activity in accordance with the intended purpose. According to Sudjana (2006), evaluation is a necessity and fairness needed in the management of a program. According to Worthen and Sanders (1981, p. 20), 'evaluation is viewed as a process of identifying and collecting information to assist decision-makers in choosing between

available decision alternatives'. Through different words, but with almost identical meanings, evaluation is described as a planned process to obtain information related to the achievement of a goal (Kartowagiran, 2013).

Evaluation is able to answer the variation of the statement and determine the success in viewing the quality of education. Rossi and Freeman (1985, p. 46) state that evaluations are conducted to answer a variety of questions related to what we have listed as the three foci of evaluation research: program conceptualization and design, program implementation, and program utility.

Weiss (1972, p. 4) writes, 'the purpose of evaluation research is to measure the effects of a program against the goals; it sets out to accomplish as a means of contributing to subsequent decision making about the program and improving future programming'. Rossi and Freeman (1985, p. 50) write that evaluation result, both from monitoring program implementation and from assessing impact and efficiency, can influence decisions on the expansion, continuation, or termination of the program and the organizations responsible for them.

This study examines the subjects of mathematics, a branch of science that has a very important role in various activities in everyday life, which can even be more than that. Thus, activities in everyday life cannot be separated from the use and application of concepts that exist in mathematics, so the unique characterization of mathematics learning is where the benefits are almost perceived in everyday life and become a key opportunity and have the contribution to other sciences.

Related to the process of its formation, mathematics is the knowledge that humans have. This knowledge arises because humans need to understand the natural world. Nature is used as a source of ideas for obtaining mathematical concepts through abstraction and idealization (Kartowagiran, 2008). If math skills can be well developed, then math can be an opportunity. This is in line with Mathematical Sciences Education Board of National Research Council (1993, p. 15) who states:

'... mathematics is the key to opportunity. No longer just the language of science, mathematics

now contributes in direct and fundamentally to business, health, and defense. For the student, it opens doors to careers. For citizens, it enables informed decisions. For nations, it provides knowledge to compute in a technological community'.

In addition, Hatfield, Edwards, Bitter, and Morrow (2008, p. 3) state that mathematics is nothing to be afraid of; it is our human heritage from all cultures. Clarifying the statement, Kahn and Kyle (2002, p. 15) explain, 'Mathematics is not fundamental too much of science and technology but needs an analytical model-building approach, whatever the discipline is'. Typically, it will be argued that mathematics claims a place in the curriculum because it can be seen as (1) contributing to the basic knowledge of any educated citizen; (2) contributing to the study and advancement of numerous disciplines, professions, and trades; (3) contributing to a student's general education through the inculcation of particular attitudes or approaches; (4) possessing an inherent interest and appeal (Christiansen, Howson, & Otte, 1986, p. 9).

The results of UNPK (*Ujian Nasional Pendidikan Kesetaraan* or National Examination of Equivalency Education) not only give results about the state of education but also provide information on improving students' learning achievement. This expectation is achieved when the data obtained are valid and reliable. In other words, the result has the smallest possible measurement error. The measurement error is divided into two: random, caused by the selection of exam materials and the condition of the examinees, and systematic, because the problem is too easy or too difficult and the implementation does not follow the guidelines, such as the regulations and operational standards of implementation (Mardapi, 2012).

The real examples of measurement error were taken from research by Kartowagiran (2008) about UAN (*Ujian Akbir Nasional* or National Examination) test device, that is, UAN Mathematics test device in 2003, 2005, and 2006 which measure three sub-dimensions of algebra, geometry, and measurement. The research found that the test devices are able to explain only 35% variance of math ability of the learners. In this regard, the test

developer should attempt to increase the factor and variance of the difficulty level of the test items.

Starting from the importance of the quality and expectations of the National Examination of Equivalency Education of Package C mathematics subjects and because the results are inseparable from the implementation process, it is really important to develop evaluation instruments for the implementation of National Examination of Equivalency Education of Package C mathematics subjects. The purpose of this study is to develop a suitable instrument for conducting the evaluation of the national examination implementation of the equivalency education Package C mathematics subject.

Method

This research and development aims to produce a particular product, and test the effectiveness of the product. The product developed is a questionnaire of the implementation of National Examination of Equivalency Education Package C consisting of 25 items. The developmental procedure referred to the modified development steps which are proposed by Mardapi (2005, pp. 16–21), as follows: (1) base on theories about the concept to be measured, as construct variables, (2) develop dimensions and indicators, (3) make instrument gratings, (4) assign quantities or parameters, (5) list instrument items, (6) validate the process, (7) revise the draft, and (8) implement the test to the Package C takers as the participants.

The evaluation instrument of National Examination of Equivalency Education Package C of mathematics subjects evaluates the standard operational procedure including the preparation, implementation, and result of the national examination. The instrument used was Likert scale modification or summative rating with the highest score per item is 4 and the lowest score per item is 1. The modified Likert scale has four options: 4 (always/strongly agree), 3 (often/agree), 2 (rarely/rather disagree), and 1 (never/disagree). The four-point scale summative rating was used because according to Mardapi (2012), in the data retrieval if using Likert scale, a five-

alternative choice such as 5 (strongly agree), 4 (agree), 3 (doubt/neutral), 2 (disagree), and 1 (strongly disagree) makes respondents often experience a tendency to choose category 3 (undecided/neutral).

The respondents in this research are Package C test participants of equality education in Bantul, Yogyakarta, Indonesia. In the try-out, the instrument was administered to 190 participants of the examination participants. The analysis of the try-out data was to obtain evidence of construct validity and reliability of the instrument. The construct validity measurement in this study used factor analysis that serves to summarize or reduce observation variables into new dimension forms that present the main variables (factors). The proof of the construct validity used exploratory factor analysis which aims to investigate the factors in the observation, and confirmatory factor analysis with the aim to confirm a theory of measurement in order to compare theories with the empirical results. The data collecting instruments with the test participants as respondents were analyzed using the exploratory factor analysis with the help of SPSS 20.0 program and followed by the confirmatory factor analysis with the help of Lisrel 8.54 program.

This research used two main factor analysis techniques: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). The CFA attempted to confirm the hypotheses and used the path analysis diagrams to represent variables and factors, and then the EFA tried to uncover complex patterns by exploring the dataset and testing predictions (Child, 2006).

The criteria in the EFA analysis must meet the following criteria: Keyser Mayer Oikin (KMO) values greater than 0.5; and the significant value of Barlett's Test of Sphericity is less than 0.05 (Ghozali, 2005). In addition, the eigenvalues of the total variances explained is greater than 1.0 and the coefficient of the Rotated Component Matrix is greater than 0.40, and the loading value of the factor is greater than that of other factors with a difference of at least 0.10 indicating a correlation between test items with a factor formed (Azwar, 2015).

Furthermore, Hendryadi and Suryani (2014) state that the criteria in the CFA analysis that can determine the suitability of the model with the help of Lisrel 8.54 can be determined as follows: (1) chi-square with p-value > 0.05; (2) Root Mean Square Error of Approximation value ≤ 0.08. Root Mean Square Error of Approximation (RMSEA) is a value that attempts to correct the trend of chi-square statistics rejecting the model; (3) the value of Goodness of Fit Index (GFI) ≥ 0.90 means that the model tested has a good match. GFI is an index that describes the overall suitability of the model of the predicted model compared to the actual data; (4) T-value ≥ 1.96 at the significance level of 0.05; and (5) Standardized loading factor > 0.5.

Furthermore, the reliability of the instrument was measured by using Alpha Cronbach formula with the help of SPSS 20.0 software and Stratified Alpha coefficient, and the reliability of the construct was measured by using the construct reliability formula. The reliability formula used in this study is as follows.

(1) Stratified Alpha Coefficient

$$\alpha_{strat} = 1 - \frac{\sum \sigma_i^2 (1 - \alpha_i)}{\sigma_x^2}$$

(2) Construct Reliability(CR)

$$CR = \frac{(\sum \text{Standardized loading})^2}{(\sum \text{Standardized loading})^2 + \sum \text{Measurement Error}}$$

(Hendryadi & Suryani, 2014)

The magnitude of the reliability index is at least 0.70 because the greater the reliability index the smaller the measurement error (Mardapi, 2012).

Findings and Discussion

The implementation of the try-out of the test is approved by using the exploratory factor analysis (EFA) with SPSS 20.0 and followed by the confirmatory factor analysis (CFA) with the help of Lisrel 8.54. through several stages of factor analysis, three times EFA and twice CFA. The steps of factor analysis to get the expected result are explained as follows.

Exploratory Factor Analysis 1

The result of the validation with exploratory factor analysis 1 on KMO value and Barlett's Test sig value is shown in Table 1.

Table 1. Exploratory factor analysis 1

Item	KMO	Sig Barlett's Test
25	0.851	0.000

Table 1 shows the value of KMO of 0.851 with a Barlett's test value of 0.000. These results show the KMO value > 0.5 and the Barlett's test < 0.05. Hence, it can be concluded that the sample size used in this factor analysis is sufficient so that the EFA analysis can proceed to the next step. Furthermore, the number of components or clusters formed from the 25 items of the statement can be seen from the total initial eigenvalues > 1.0 shown in Table 2.

Table 2. Eigenvalues 1

Component	Initial Eigenvalues		
	Total	Variance (%)	Cumulative (%)
1	6.780	27.119	27.119
2	2.538	10.154	37.273
3	1.900	7.600	44.873
4	1.730	6.921	51.794
5	1.101	4.405	56.199

Table 2 shows that the total initial eigenvalues > 1.0. Thus it can be concluded that there are 5 components formed from 25 items in the instrument with the variance described as 56.199%. Furthermore, the number of factors in the instrument can be seen in the scree plot shown in Figure 1.

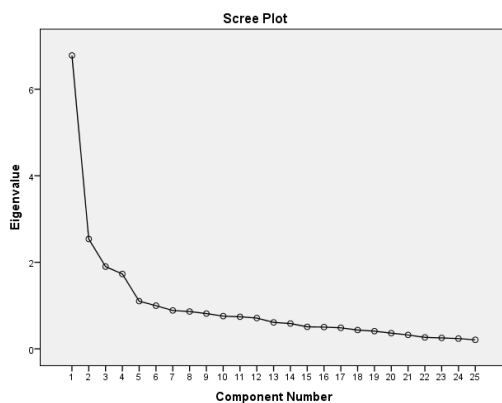


Figure 1. Scree plot EFA 1

Figure 1 shows the number of factors marked by a steep graph of eigenvalue value gain. Based on Figure 1 then, there is one dominant factor and the other four factors also contributing substantially to the component of the variance that can be explained, so that the instrument shows it measures at least five factors that are formed. Thus, it can be concluded that all of the items can be analyzed further through factor analysis with the extraction and rotation method using varimax and obtained the results as shown in Table 3.

Table 3. Rotated component matrix 1

	Component				
	1	2	3	4	5
A1	0.637				
A2	0.709				
A3					
A4	0.772				
A5	0.788				
A6	0.661				
A7	0.465				
A8	0.697				
B1		0.725			
B2		0.798			
B3		0.458			
B4		0.756			
B5		0.578			
B6		0.620			
B7					
B8					0.736
C1			0.442		
C2			0.752		
C3			0.820		
C4					
C5			0.634		
D1				0.766	
D2				0.823	
D3				0.803	
D4				0.590	

Table 3 shows the value of the loading factor does not meet the specified criterion, that is ≥ 0.4 and the difference with another factor > 0.1. There are three invalid items: A3 (item 3) which forms learning time factor, B7 (point 15) which forms socialization formation factor, and C4 (item 20) which forms examination material factor. This can happen because of the different interpretation between the researchers and respondents. Because the items are not good to use, then the invalid items are discarded then proceed with second exploratory analysis with 22 items.

Exploratory Factor Analysis 2

The second exploratory factor analysis was performed after the invalid items were discarded. Thus, there were 22 items left to be analyzed.

Table 4. Exploratory factor analysis 2

Item	KMO	Sig Barlett's Test
22	0.838	0.000

Table 4 shows the gain of KMO value of 0.838 with a Barlett's test value of 0.000. This result shows $KMO > 0.5$ and Barlett's test < 0.05 . It can be concluded that the sample used in this research is adequate. Furthermore, the number of components or clusters formed from the 22 items can be seen from the total initial eigenvalues > 1.0 .

Table 5. Eigenvalues 2

Component	Initial Eigenvalues		
	Total	Variance (%)	Cumulative (%)
1	6.216	28.254	28.254
2	2.446	11.119	39.372
3	1.718	7.811	47.183
4	1.592	7.235	54.419

Table 5 shows the total initial eigenvalues is > 1.0 . It can be concluded that there are four clusters formed from 22 items on the sheet of an instrument with a cumulative percentage of 54.419% and it explains the variance. Furthermore, the scree plot also shows there are four dots that are above the value of 1 because the number of factors is marked by a steep graph of eigenvalue value gain, then there is one dominant factor and three other factors also contributing substantially to the cluster variance that can be explained so that the instrument measured at least four factors and clarified on the scree plot as in Figure 2

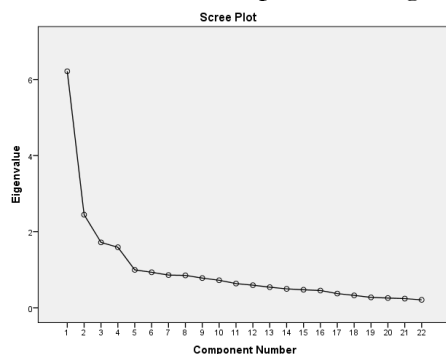


Figure 2. Scree plot EFA 2

Thus, it can be concluded that the whole items can be analyzed further through the factor analysis with the extraction and rotation method using varimax and it obtained the results as in Table 6.

Table 6. Rotated component matrix 2

	Component			
	1	2	3	4
A1	0.646			
A2	0.725			
A4	0.773			
A5	0.807			
A6	0.657			
A7	0.457			
A8	0.698			
B1		0.730		
B2		0.780		
B3				
B4		0.767		
B5		0.597		
B6		0.664		
B8				
C1				0.474
C2				0.801
C3				0.849
C5				0.615
D1			0.760	
D2			0.820	
D3			0.803	
D4			0.596	

Table 6 shows that the value of the loading factor does not meet the critereon specified that is ≥ 0.4 and difference with other factors > 0.1 . There are two invalid items, namely B3 (point 11) and B8 (point 16), the items which form the socialization factor. The invalid items may be caused by the difference of interpretation between the researchers and the respondents or the items are unfavorable to use. Thus, the invalid items are discarded and then followed by the third exploratory factor analysis with 20 items.

Exploratory Factor Analysis 3

This third exploratory factor analysis was performed after invalidating the invalid items. Therefore, 20 items can be analyzed.

Table 7. Exploratory factor analysis 3

Item	KMO	Sig Barlett's Test
20	0.834	0.000

Based on Table 7, KMO value is 0.834 with Barlett's test significance value of 0.000.

These results show that the KMO value is > 0.5 and value of Barlett's test is < 0.05 . It can be concluded that all items in the instrument can be analyzed further. Furthermore, the number of components or clusters formed by the 20 items of the statement can be seen from the total initial eigenvalues > 1.0 .

Table 8. Eigenvalue 3

Comp	Initial Eigenvalues		
	Total	Variance (%)	Cumulative (%)
1	5.804	29.020	29.020
2	2.410	12.052	41.071
3	1.680	8.398	49.469
4	1.552	7.760	57.230

Table 8 shows the total initial eigenvalues is > 1.0 . Thus, it can be concluded that there are four components formed by 20 items in the instrument with a percentage value of 57.230% variance that can be explained. A model that is a good fit will have less than 50% of the non-redundant residuals with absolute values that are greater than 0.05 (Yong & Pearce, 2013). Furthermore, the number of factors in the instrument can be seen through the scree plot shown in Figure 3.

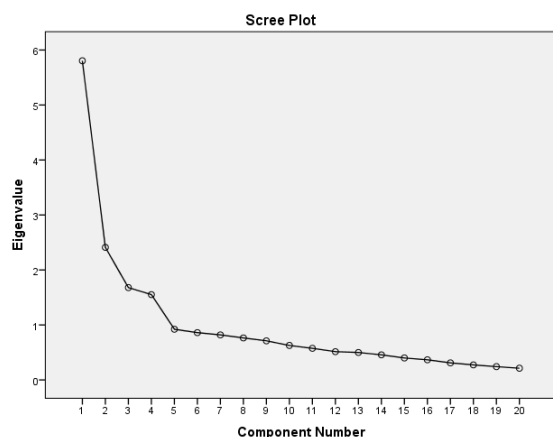


Figure 3. Scree plot EFA 3

Figure 3 shows that the number of factors is marked by a steep graph of eigenvalue gain. Based on the figure, there is one dominant factor and the other three factors also contributing substantially to the component of variance that can be explained and they begin to ramp up on a fifth factor. This indicates that the instrument shows at least four factors. The scree test consists of eigenvalues

and factors (Cattell, 1978). The scree test is only reliable when the sample size is at least 200. In situation when the scree test is hard to interpret, it is necessary to rerun the analysis several times and manually set the number of factors to extract each time (Costello & Osborne, 2005).

Thus, it can be concluded that whole items can be analyzed further through factor analysis with the extraction and rotation method using varimax aiming to clarify the items included in the component. Yong and Pearce (2013) write that factors are rotated for better interpretation, since unrotated factors are ambiguous. Thus, the results obtained can be seen in Table 9.

Table 9. Rotated component matrix 3

	Component			
	1	2	3	4
A1	0.650			
A2	0.724			
A4	0.770			
A5	0.804			
A6	0.660			
A7	0.463			
A8	0.697			
B1		0.740		
B2		0.787		
B4		0.766		
B5		0.596		
B6		0.656		
C1				0.505
C2				0.810
C3				0.849
C5				0.622
D1			0.743	
D2			0.827	
D3			0.806	
D4			0.607	

Based on the EFA analysis, the four components formed by the 20 items are elaborated as follows: (1) The items related to group learning time are clustered in component 1; (2) the items related to socialization are clustered in component 2; (3) the items related to examination material are clustered in component 4; and (4) the items associated with the examination room are grouped in component 3. On the other hand, the clusters obtained by the exploratory factor analysis are then analyzed by using the Confirmatory Factor Analysis (CFA).

Confirmatory Factor Analysis

Before calculating the validity of the construct by using the CFA, the assumption of normal distribution is firstly tested. This normality test can be seen from the univariate normality test result that describes the distribution of one variable in the respondent, whereas the multivariate normality test result provides an overview of the shared distribution of all variables in the respondent. The calculations in this study employed Lisrel 8.54 program.

The results of the univariate normality analysis are shown in Table 10. Furthermore, the summary results of multivariate normality calculations are shown in Table 11.

The results of the univariate normality analysis showed that the data did not meet the normal univariate assumptions (p value skewness and kurtosis <0.05), in line with the results of the normal multivariate test which was not fulfilled (p value skewness and kurtosis

<0.05). It can be concluded that the data used do not meet normal univariate or multivariate assumptions. Normal univariate distribution of each item is required, but multivariate distribution is more important because in general, data with no normal univariate distribution will result in a multivariate non-normal distribution (Hendryadi & Suryani, 2014).

Furthermore, due to the abnormal data, this research used an alternative estimation method that is Robust Maximum Likelihood (RML) by adding asymptotic covariance matrix which is useful for correcting the chi-square statistic value, commonly known as Satorra-Bentler Scaled Chi-Square. Maximum Likelihood attempts to analyze the maximum likelihood of sampling the observed correlation matrix (Tabachnick & Fidell, 2007). The Maximum Likelihood is more useful for confirmatory factor analysis (Yong & Pearce, 2013).

Table 10. Test of Univariate Normality for Continuous Variables

Variable	Skewness		Kurtosis		Skewness and Kurtosis	
	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
A1	1.917	0.055	-1.281	0.200	5.316	0.070
A2	-0.126	0.900	-1.711	0.087	2.945	0.229
A4	-3.319	0.001	-0.690	0.490	11.492	0.003
A5	0.608	0.543	-1.932	0.053	4.101	0.129
A6	-1.237	0.216	-2.507	0.012	7.817	0.020
A7	-1.785	0.074	-0.262	0.793	3.254	0.197
A8	0.732	0.469	-3.458	0.001	12.478	0.002
B1	-2.578	0.010	-1.498	0.134	8,890	0.012
B2	-3.566	0.000	0.749	0.454	13.276	0.001
B4	-3.673	0.000	1.550	0.121	15.892	0.000
B5	-4.506	0.000	2.549	0.011	26.808	0.000
B6	-2.866	0.004	1.552	0.121	10.626	0.005
C1	-3.338	0.001	2.086	0.037	15.493	0.000
C2	-5.979	0.000	2.360	0.018	41.311	0.000
C3	-5.837	0.000	2.415	0.016	39.907	0.000
C5	-2.525	0.012	1.634	0.102	9,047	0.011
D1	-3.361	0.001	2.243	0.025	16.322	0.000
D2	-2.063	0.039	0.388	0.698	4.407	0.110
D3	-1.836	0.066	0.238	0.812	3.428	0.180
D4	-4.165	0.000	-0.559	0.576	17.660	0.000

Table 11. Test of multivariate normality for continuous variables

Value	Skewness		Value	Kurtosis		Skewness and Kurtosis	
	Z-Score	P-Value		Z-Score	P-Value	Chi-Square	P-Value
8.177	18.276	0.000	518.921	10.858	0.000	451.914	0.000

The CFA (Confirmatory Factor Analysis) was based on an exploratory analysis which resulted in four components and was supported by the existence of theory, and then was subsequently analyzed by the confirmatory analysis. The calculation of CFA was done twice with the help of Lisrel 8.54 program. The first result was Root Mean Square Residual (RMR) = 0.0346, Goodness of Fit Index (GFI) = 0.881 and Root Mean Square Error of Approximation (RMSEA) = 0.0421 and Satorra-Bentler Scaled Chi-Square = 281.932 with P-value 0.00268. The standardized solution model of the Package C execution of mathematics subjects is clearly presented in Figure 4, while when it is seen from t-value, the result is shown in Figure 5.

When viewed from the results and models obtained, P-value $0.00268 < 0.05$, this indicated that the factor model used by all the tests was not good (the model was not fit). Therefore, to get the fit model, model respecification or model modification by looking at the modification indices to see the items that correlate each other was done. The results of the second-factor analysis calculation, after getting the correlated items, was gained through using modification indices as a reference. Thus, the standardized solution model 2 of the implementation of the Package C mathematics subject through Confirmatory Factor Analysis 2 is presented in Figure 6, while when it is seen from t-value, the result is shown in Figure 7.

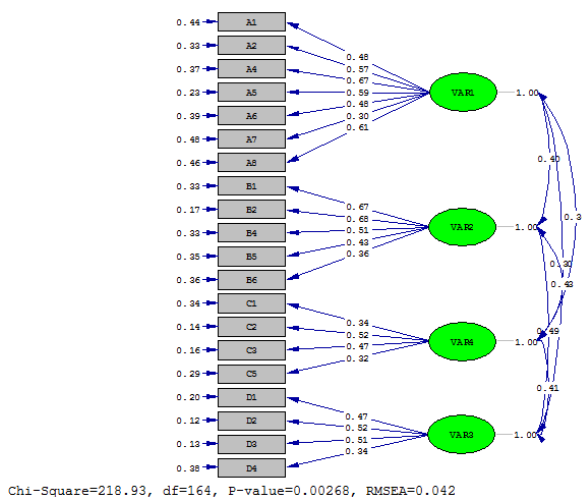


Figure 4. Standardized solution 1

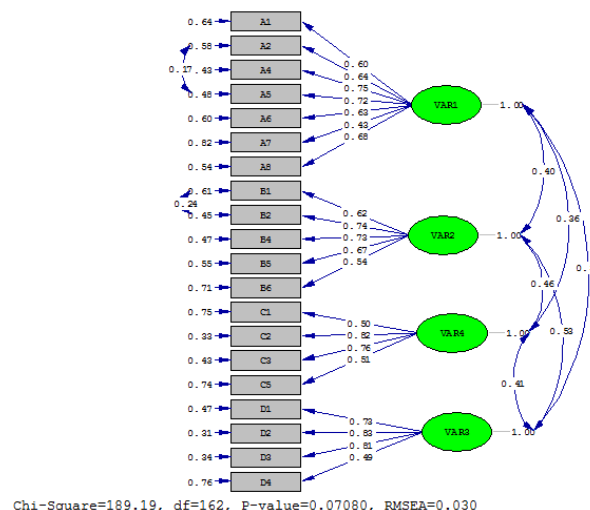


Figure 6. Standardized solution 2

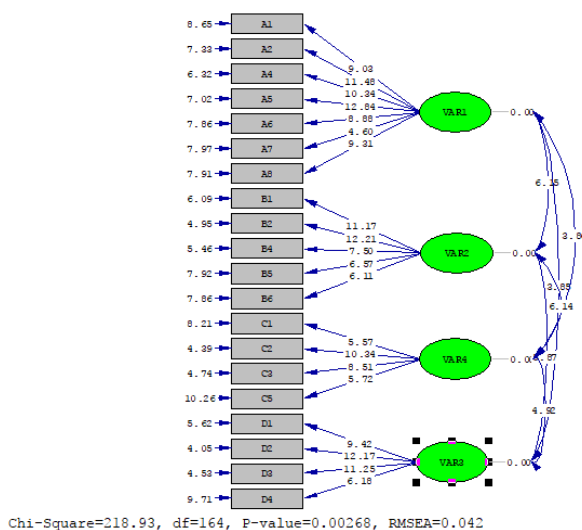


Figure 5. T-value 1

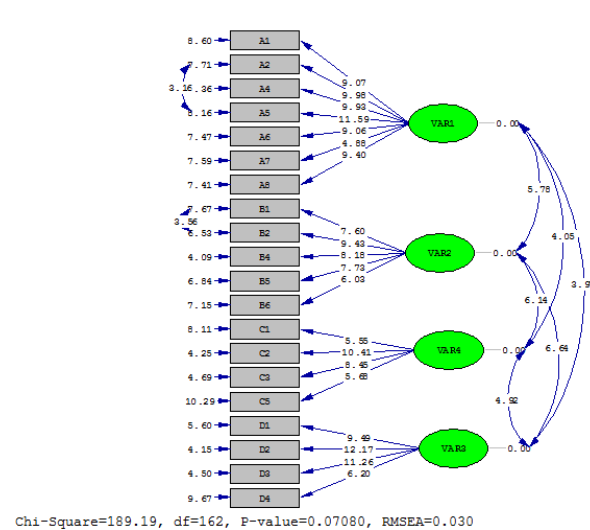


Figure 7. T-value 2

The results of this CFA 2 output show that the value of Goodness of Fit Index (GFI) = 0.894. and Root Mean Square Error of Approximation (RMSEA) = 0.030 < 0.080 (good fit) and Satorra-Bentler Scaled Chi-Square = 189.186 with P-value 0.0708 > 0.050 (good fit). Seen from the results and match models, the proposed model has a good match or the proposed model matches the data and the conceived items measure only the latent variables. The correlated items are due to identical statements. The result of the weighted coefficient significance of the 20 items rated on Standardized Loading Factors (SLF) shows that two items have less than 0.5, i.e. items A7 and D4. However, the overall value of t-value is > 1.96. Thus, there are two items with poor validity: items A7 and D4. Thus, based on the Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA), it can be concluded that the instrument of the National Examination of Equivalency Education Package C is valid to measure the implementation and it is proven to be empirical.

Furthermore, the reliability of the instrument, in which the respondents take the role as the test participants, was calculated by using alpha Cronbach formula. Stratified Alpha and Construct Reliability (CR) were used to determine the reliability of the constructs. The component and total reliability coefficients were sought. The reliability coefficient results are shown in Table 12.

Based on Table 12, it can be concluded that the instrument with the test participants as the sample is stated to be reliable, and thus the instrument with 20 items can be valid and reliable if it is re-measured by using the same object because it has quite high reliability and feasibility value. It has the reliability coefficient value of at least > 0.7. It states that the used indicators already have adequate internal

consistency reliability, meticulous in measuring and explaining the construct.

After that, some steps conducted during the research produced a final product, which was used as a questionnaire instrument developed from the Standard Operational Procedure (SOP) of the national examination. From the try-out of 25 items, only 20 items fulfilled the standard validity and reliability, so it was found using removal information. The result shows that it has validity value > 0.40 and reliability coefficient value > 0.70. Overall, the results show that the developed instrument is equal with the SOP of the national examination and it has been proven empirically that it is in a good category.

Conclusion and Suggestions

The research concludes that (1) based on the tested data with the test takers as the respondents, the instrument is valid, reliable, and qualified as the fit model; (2) the components in the instrument are the learning time, socialization, test materials, and examination room; and (3) the instrument has the validity value of > 0.40 and reliability value of > 0.70.

Based on the findings, some suggestions are proposed as follows: (1) the instruments developed in this model were only applied to the test takers as respondents. Thus, it is suggested to other researchers to develop it further, so that the evaluation instrument of national examination implementation of Package C at equality education will be better; and (2) the coverage of the objects in the evaluation instrument of the National Examination of Equivalency Education implementation of Package C is still too narrow, and therefore, other researchers need to add other components of the implementation so that the coverage can be more comprehensive.

Table 12. The reliability coefficient results of UNPK Evaluation

No	Component	Cronbach's Alpha coefficienty	Stratified Alpha coefficienty	CR	Remark
1	Learning time	0.831	-	0.886	Reliable
2	Socialization	0.809	-	0.865	Reliable
3	Examination material	0.734	-	0.826	Reliable
4	Examination room	0.797	-	0.878	Reliable
	Total	-	0.902	0.870	Reliable

References

- Adow, I. M., Alio, A. A., & Thinguri, R. (2015). An assessment of the management of KCSE examination and its influence on irregularities among students: A case of secondary schools in Mandera County, Kenya. *Journal of Education and Practice*, 6(28), 15–22.
- Azwar, S. (2015). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum Press.
- Child, D. (2006). *The essentials of factor analysis* (4th ed.). New York, NY: Continuum.
- Christiansen, B., Howson, A. G., & Otte, M. (Eds.). (1986). *Perspectives on mathematics education*. Dordrecht: Springer Netherlands.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1–9. Retrieved from <http://www.statsoft.com/textbook/>
- Ghozali, I. (2005). *Structural equation modeling: Teori, konsep, dan aplikasi dengan program Lisrel 8.80*. Semarang: Badan Penerbit Universitas Diponegoro.
- Government Regulation No. 19 of 2005, on National Education Standard (2005). Republic of Indonesia.
- Hatfield, M. M., Edwards, N. T., Bitter, G. G., & Morrow, J. (2008). *Mathematics methods for elementary and middle school teachers*. Hoboken, NJ: John Wiley and Sons.
- Hendryadi, & Suryani. (2014). *Structural equation modeling dengan Lisrel 8.80*. Jakarta: Kaukaba Dipantara.
- Kahn, P., & Kyle, J. (Eds.). (2002). *Effective learning and teaching in mathematics and its applications*. London: Routledge.
- Kartowagiran, B. (2008). Validasi dimensionalitas perangkat tes ujian akhir nasional SMP mata pelajaran matematika 2003-2006. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 12(2), 177–195. <https://doi.org/10.21831/pep.v12i2.1426>
- Kartowagiran, B. (2013). *Evaluasi dan pengembangan kurikulum*. A paper presented in Workshop Evaluasi Kurikulum STABN Raden Wijaya.
- Law No. 20 of 2003 of Republic of Indonesia on National Education System (2003).
- Mardapi, D. (2005). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.
- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Mardapi, D., & Kartowagiran, B. (2009). *Dampak ujian nasional*. A research report. Yogyakarta: Program Pascasarjana UNY.
- Mathematical Sciences Education Board of National Research Council. (1993). *Measuring what counts: A conceptual guide for mathematics assessment*. Washington, DC: National Academy Press.
- Mudjijanti, F. (2011). Pengaruh tes masuk berdasarkan nilai ujian nasional (UN) terhadap prestasi belajar siswa (Studi kasus di SMUK St. Bonaventura Madiun). *Widya Warta*, 35(2), 19–31.
- Raharjo, S. B. (2012). Evaluasi trend kualitas pendidikan di Indonesia. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 16(2), 511–532. <https://doi.org/10.21831/pep.v16i2.1129>
- Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: A systematic approach*. Beverly Hills: Sage.
- Setiadi, H. (2011). *Analisis kelemahan kompetensi siswa pada tingkat kabupaten/kota berdasarkan hasil UN rendah tahun 2011 di Kabupaten Alor*. A research report. Jakarta: Badan Penelitian dan Pengembangan Pusat Penilaian Pendidikan.

- Sudjana, D. (2006). *Evaluasi program pendidikan luar sekolah untuk pendidikan nonformal dan pengembangan sumber daya manusia*. Bandung: Remaja Rosdakarya.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (4th ed.). Hillsdale: Erlbaum.
- Weiss, C. H. (1972). *Evaluation research*. London: Prentice-Hall.
- Worthen, B. R., & Sanders, J. R. (1981). *Educational evaluation: Theory and practice*. Worthington, OH: Charles A. Jones.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), 79–94. <https://doi.org/10.20982/tqmp.09.2.p079>