THE ACCURACY OF THE CHEATING DETECTION METHODS
IN LARGE-SCALE TESTS: MATHEMATICS NATIONAL EXAMINATION

Thomas Mbenu Nulangi [1] **, Djemari Mardapi* [1]
[1]Universitas Negeri Yogyakarta
[1]Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
* Corresponding Author. Email: tommymbenu07@gmail.com

Abstract

This study aimed to describe (1) the characteristics of items based on the Item Response Theory, (2) the cheating level in the implementation of the national examinartion based on Angoffs B-Index method, Pair 1 method, Pair 2 method, Modified Error Similarity Analysis (MESA) method, and G2 method, (3) the most accurate method to detect the cheating in the mathematics national examination at the senior secondary school level in the academic year of 2015/2016 in East Nusa Tenggara Province. The result of the item response theory analysis showed that 17 (42.5%) items of the mathematics national examination fit with the 3-PL model, with the maximum information function of 58.0128 at $\theta =$1.6, and the measurement error of 0.1313. The number of pairs detected to be cheating by Angoff's B-Index method was 63 pairs, that by the Pair 1 method was 52 pairs, that by the Pair 2 method was 141 pairs, that by MESA method was 67 pairs, and that by the G2 method was 183 pairs. The methods which could detect most pairs doing cheating were the G2 method, the Pair 2 method, the MESA method, Angoff's B-Index method, and the Pair 1 method successively. The methods which could accurately detect cheating based on the computation of the standard error were Angoff's B-Index method, the G2 method, the MESA method, the Pair 1 method, and the Pair 2 method successively.

Keywords: *national examination, item characteristics, cheating detection methods*

Introduction

The administration of the national examination as regulated in Government Regulation No 13 Year 2015 (Presiden Republik Indonesia, 2015) about the Educational National Standard aims to reveal how far the students have already mastered the competence. This is in accordance with the competence-based instructional objectives that students should know the development of their learning ability. The competence-based teaching should be implemented to obtain information about students' success and teachers' success in teaching. In the competence-based teaching, it is necessary to establish the graduate competence standard which include attitudes, knowledge, skills so that they can be a reference for schools to implement the teaching activities.

The Educational National Standard Board (*Badan Standar Nasional Pendidikan/* BSNP) as an independent institution is responsible for administering the national examination objectively, fairly, and accountably so that the result of the national examination can provide information about the students' learning success and teachers' success in teaching (Mulyati & Kartowagiran, 2013). Therefore, honesty and achievement become the motto of the administration of the national examination in 2015 (Badan Standar Nasional Pendidikan, 2015, p. 4) The 2015 national examination motto should be considered as the common principle which should be owned by all stakeholders in the administration of the national examination so that the national examination administration is of high quality.

Up to the moment, the Indonesian educational system still uses the final achievement as the indicator which the progress and the students' mastery of knowledge. As a result, the community tends to view learning achievement as only seen from the high final scores not in the learning process. This view has caused burden to students to get as high scores as possible (Manoppo & Mardapi, 2014). The burden will make students be oriented to getting high scores, not to mastering knowledge so that students can do anything including cheating in the examination.

Cheating behaviour is not expected to happen in the administration of the broad scale testing like the national examination. Cheating behaviour is considered as violating the law because this will give benefit to those doing cheating. In line with this, Cizek (2001, p. 7) states that cheating can be defined as behaviour which violates the law for administering a test, every behaviour which gives benefits unfairly in the examination, or the behaviour of the test administrator which can decrease the accuracy of the test scores or the examinees' performance.

Cheating behaviour can be found in all levels and done systematically. As stated by Davis, Drinan, & Gallant (2009, p. 1), students of all educational levels do cheating, from the primary schools to the graduate program, from schools located in villages to those located in cities, from poor schools to rich schools, both public and private schools. Students do cheating because they are afraid that they will get poor scores. They are not honest to themselves. They are not honest their friends and their parents. They sometimes even work hand in hand with teachers and school staff to do cheating.

The problem about cheating is a classical problem faced by all nations. Findings of questionnaires and interviews by Mardapi (2000) when doing an evaluation of the administration of the national examination in 2000 showed that 76% grade I senior secondary school students in Central Java, 73.9% grade I senior secondary school students in South Sulawesi, and 81.8% senior secondary school students in Jambi did cheating in the national examination. The cheating could be done in the form of opening the notes, peeping at others' work, or doing something not allowed according to the regulation. Further findings showed that 20% grade I senior secondary school students in Cental Java, 26.1% grade I senior secondary school students in South Sulawesi, and 18.2% grade I senior secondary school students in Jambi said that the proctors assisted the students in doing the exami-

nation. This happened because the proctors were kind so that the standard of the national examination administration had not been implemented in all schools. Another problem was related to the true examination scores (*Nilai Ebtanas Murni*/NEM). There were 12.6% teachers in Central Java, 23.8% teachers in South Sulawesi, and 46.8% teachers in Jambi thought that the true examination scores (NEM) were not true scores in the real sense.

The Ministry of Education and Culture through its website (http://www.kemdikbud.go.id) presented news that the national examination integrity index in 2015 was 64.05 nationally. This means that cheating was still done in the administration of the national examination in Indonesia in 2016. In addition, the information provided by the ministry of education and culture in 2015 showed that the national examination integrity index of East Nusa Tenggara province was 62.2%. In 2016 the national examination integrity index of East Nusa Tenggara was 69.3%. Although there was an increase, there were still 30.7% considered as doing cheating in the national examination (Biro Komunikasi dan Layanan Masyarakat Kementrian Pendidikan dan Kebudayaan, 2016).

Cheating behaviour does not only happen in Indonesia but also in other countries such as America. Naghdipour & Emeagwali (2013) conducted a study on 500 consisting of 450 students and 50 lecturers from different faculties in America. The findings showed that there was an indication of academic dishonesty, although students did not make any report on this. In addition, there was a difference between the report as provided by the students and that provided by lecturers in their observation about the number of students doing cheating.

The cheating behaviour will give an impact on the students' future if this is done again and again. The study conducted by Bernardi, Banzhoff, Martino, & Savasta, (2012) showed that the cheating behaviour done again and again from early childhood to adulthood will result in social problems.

The cheating behaviour will influence students' willingness to do the same in their social life in the future. The study conducted by Zastrow (1970) on 45 graduate students showed that 40% of the students did cheating. The reason for doing cheating is the burden from the parents. The parents expected them to get good marks. In addition, this study also found out that there was no difference in the characteristics between students who did cheating and those who did not. This means that the cheating behaviour was done systematically.

There are some ways of doing cheating in the examination. The study conducted by Baird (1980) on 200 students showed that 75% of the students did cheating in the examination. Baird found twelve ways of doing cheating, such as, getting information from other students, copying others' work, copying others' assignment, copying from books, copying others' work, hiding errors from the lecturers, illegal test information, stealing others' test instruments, changing the examination paper, taking a test for others, others taking test for a student, bribery and blackmailing in an examination.

Previously, there were some studies focusing on demographic aspects such as gender Anderman & Murdock (2007, p. 11). The study focused on the difference in gender in doing cheating. The study revealed that male students tended to do more cheating than female students (Calabrese, & Cochran, 1990; Davis, Grover, Becker, & McGregor, 1992; Michaels & Miethe, 1989; Newstead, Franklyn-Stokes, & Armstead, 1996).

However, Whitley, Nelson, & Jones (1999) Nelson and Jones (1999) revealed the effect of the attitude size effect of gender in the medium size and the effect was so small so that they all did cheating. Therefore, both male and female students had the tendency to do cheating.

Doing cheating in the class must be stopped by cutting the missing link. Therefore, teachers should be more proactive in stopping the cheating behaviour in the class. Some simple ways to decrease cheating in the

class such as distributing different test forms, making the seating arrangement randomly, improving he supervision when proctoring an examination, and making spaces between seats (Bernardi, Baca, Landers, & Witek, 2008). Those ways are considered as being effective in reducing the cheating behaviour in the administration of the national examination.

The advancement of technology has contributed greatly to cheating, called e-cheat. A survey conducted by Stogner, Miller, & Marcum (2013) on 534 students in universities showed that 40% of the students did e-cheating the year before. They did cheating using e-cheating because it was free of charge. The findings of the study suggested that teaching restrict cheating behaviour.

Many survey reports showed that copying answers was very common cheating behaviour (Bopp, Gleason, & Misicka, 2001; Brimble & Clarke, 2005; Hughes & McCabe, 2006; Jensen, Arnett, Feldman, & Cauffman, 2002; Lin & Wen, 2007; Rakovski & Levy, 2007; Vandehey, Diekhoff, & LaBeff, 2007). Copying answers could result in uncommon response patterns for the students doing the copying or uncommon answer similarity between two responses, which can be studied through the probabilistic model.

Some statistical procedures have been reported in literature for the last few years related to copying answers. To detect the act of copying answers, some methods are developed, such as Baird Index, Dickenson Index, Anikeef Index (Anderman & Murdock, 2007, pp. 263–264), A, B, and H Index developed by Angoff (1972), Pair 1 method and Pair 2 method developed by Ferry, Tidman, & Wats (1977), G2 method developed by Hanson, Haris, & Brenan (1987), *Error Similarity Analisys Index (ESA)* and *Modified Error Similarity Analisys (MESA)* developed by Bellezza dan Bellezza (Anderman & Murdock, 2007, p. 267) and some other methods to detect the act of cheating.

Based on the above information, what is meant by cheating is every form of dishonest behavior in cheating done by a person to gain success in doing the academic assignment, especially related to examination. The above facts showed that cheating was an interesting problem to be analyzed further. Therefore, through this study, the researcher examined some methods to detect the possibility of the students doing cheating or sometimes called collusion through observing the response patterns from the examinees.

This study aimed to describe: (1) the characteristics of the mathematics national examination items based on Item Response Theory, (2) the cheating level happening in the administration of the mathematics national examination for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province based on Angoffs B-Index, the Pair 1 method, the Pair 2 method, Modified Error Similarity Analysis (MESA), and the G2 method, and (3) the most accurate method to detect the act of cheating in the administration of the mathematics national examination for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province.

Method

This study belongs to quantitative research using the ex-post facto approach. This study was carried out in East Nusa Tenggara Province. The data analysis was done in the computer laboratory of the Graduate Program, Yogyakarta State University after obtaining the sample response data from *Puspendik, Kemendikbud*, Jakarta. This study was carried out in March 2017.

The target of this study was the senior secondary school mathematics test in the academic year of 2015/2016 Package 3324. The data used in this study were the students' responses in the national examination in East Nusa Tenggara Province, especially on mathematics involving 3,233 students.

The procedures of the study were as follows: (1) estimating the fitness of the data and the model used using the software Bilog MG version 3.0; (2) analysing the items using the Item Response Theory to find out the

item parameter, the examinee parameter, the test information function (the scale taken was in the range between -4.0 and 4.0 with an interval of 0.25) and the standard error of measaurement, (3) detecting the act of cheating using five methods, i.e. the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA), and the G2 method with the help of *Integrity* software, (4) determining the method which could detect most pairs suspected to have done cheating, (5) determining the accuracy of the five methods in detecting the act of cheating through the standard error.

The data used in this study were secondary data in the form of examinees' responses on mathematics national examination for senior high schools in East Nusa Tenggara obtained from the *Puspendik, Kemendikbud*, Jakarta. The number of responses was 3,233 students doing package 3324. The data were raw data ABCDE.

The data collection technique was documentation. The data analysis technique was the quantitative data analysis technique. There were two stages in the data analysis, i.e. the item analysis using the Item Response Theory on the senior secondary school mathematics national examination in the academic year of 2015/2016 and the analysis of cheating using the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA) method, and the G2 method.

The analysis of the test item in this study used the Item Response Theory using the Bilog-MG program. This analysis resulted in item information in accordance with the Item Response Theory used. Then, using the 1 Logistic Parameter, the estimation of the difficulty index was obtained. Using the 2 Logistic Parameter, the difficulty index and discrimination index were obtained. Using the 3 Logistic Parameter, the information on the difficulty index, discrimination index, and the artificial guessing. (DeMars, 2010, p. 34) argued that for 1 Logistic Parameter model, as many as 100 or 200 students could be used. For 2 Logistic Parameter model and 3 Logistic Parameter model, there must be more than 500 students used.

In addition to the item parameter, the goodness of fit statistics was also obtained. The model used for estimating the parameter was the logistic model which could accept many item fits. An item which does not fit is the one with a chi-square value higher than the chi-square critical value. On the other hand, an item which fits is the one with a chi-square value lower than the chi-square critical value or the probability value is more than 0.05.

In addition, there was also another information obtained in this analysis, that is, the test information function and the standard error of measurement (SEM). The value of the test information function was calculated using the excel program. The above criteria were used to determine the good item quality using the Item Response Theory. The criteria were as follows: (1) the discrimination index was in the range between 0 and 2, and (2) the difficulty index was in the range between -2 and +2, and (3) the guessing index was around 0.5.

The analysis of the act of cheating in this study used the help of *Integrity*[TM] software. This software is a safe online application which is designed to analysis multiple choice examination. This software uses five methods, that is, the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA) method, and the G2 method. There were two packages which can be utilized by clients in analysing the act of cheating, i.e. using the free trial package and the purchase license model. This study used the free trial package.

Findings and Discussion

The Item Characteristics based on the Item Response Theory

The test item empirical analysis in this study used the Item Response Theory approach using the Bilog-MG software. Before the item analysis was done, a test on the model fitness was done. It was aimed to find

out the model which fitted with the students' response characteristics.

To know the fitness of the model, a criterion of having the probability value of >0.05 was used. The result of the model fitness test can be seen in Table 1.

Table 1. The Item Fitness to *1-PL*, *2-PL*, *3-PL models*

Parameter Logic	The model fit item	Total
1 PL	12, 29	2
2 PL	12, 23, 24, 31	4
3 PL	5, 12, 23, 25, 29	5

The test of model fitness showed that the senior secondary school mathematics examination in East Nusa Tenggara Province in the academic year of 2015/2016 for package 3324 fitted with the 3 Logistic Parameter model. This is because the 3 Logistic Parameter model could detect most items fitting the model. The number of the items fitting the 3 Logistic Paremeter model was 5 items so that the item parameter characteristics analysis used the 3 Logistic Parameter model. The five items were item number 5, 12, 23, 25 and 29.

The next analysis was finding out the item characteristics using the 3 Logistic Parameter model. The good difficulty index criteria were in the range between -2 and 2. When an item has a difficulty index lower than -2, this item is considered as a difficult item, while if the difficulty index is more than 2, the item is considered as easy.

The item characteristics of the mathematics national examination package 3324 viewed from the difficulty level was that there were 33 (82.5%) items belonging to the difficulty at the medium level. The rest or 7 items (17.5%) belonged to the difficult items. The seven items were item no 4, 16, 21, 24, 30, 31, 33, dan 36. However, there were many items belonging to the difficulty of the medium level. Meanwhile, the mean of the difficulty level was 2.565. Therefore, it could be concluded that the mathematics national examination package 3324 for senior sec-

ondary school students in East Nusa Tenggara Province in the academic year of 2015/2016 belonged to the difficult test. This might cause students with low ability to find difficulty to do the test. This means that there were seven items which could only be answered correctly by those with high ability.

Table 2. The Item Difficulty Index of the Mathematics National Examination Package 3324

Difficulty level	Item no	Total
Easy	-	-
Medium	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 25, 26, 27, 28, 29, 32, 34, 35, 37, 38, 39, 40	31
Difficult	4, 16, 21, 23, 24, 30, 31, 33, 36	9

The criteria related to the discrimination index are that the item discrimination index was in the range of 0 and 2. An item having a discrimination index of more than 2 indicated that the item had a good discrimination index. Meanwhile, a lower discrimination index showed contrary information. The result of the discrimination index is presented in Table 3.

Table 3. Mathematics National Examination Package 3324 Discrimination Index

Discrimination index	Item no	Total
Good	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 18, 19, 21, 23, 24, 25, 26, 29, 30, 31, 32, 33, 35, 36, 38	30
Poor	11, 17, 20, 22, 27, 28, 34, 37, 39, 40.	10

Viewed from the characteristics of discrimination index parameter, there were 30 (75%) items belonging to the good category. The rest (10 items or 25%) belonged to

the poor category. The ten items were item no 11, 17, 20, 22, 27, 28, 34, 37, 39, and 40. The mean of the discrimination index was 1.543, which means that the mathematics national examination package 3324 for the senior secondary school students in East Nusa Tenggara Province in the academic year of 2015/2016 had good discrimination index. This means that the test could distinguish students with high and low abilities, while ten items could not distinguish students with high and low abilities.

The criteria for the artificial guessing were in the range between 0 and 0.20. A high artificial guessing index shows that it is highly possible that the examinee guesses the answer correctly while a low artificial guessing index indicates that there is a small possibility that the examinee guesses the answer correctly. The result of the analysis of the artificial guessing can be seen in Table 4.

Table 4. Artificial Guessing of the Mathematics National Examination Package 3324

Artificial guessing	Item no	Total
Good	1, 2, 3, 4, 5, 66, 8, 9, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 26, 27, 29, 30, 31, 32, 33, 35, 36, 38, 40	30
Not good	6, 10, 11, 12, 22, 25, 28, 34, 37, 39	10

Viewed from the parameter of the artificial guessing, there were 30 (75%) items belonging to the good category. The rest (10 items or 25%) belonged to the poor category, i.e. item no 6, 10, 11, 12, 22, 25, 28, 34, 37, and 39. The mean of the artificial guessing index was 0.135. Viewed from the mean of the artificial guessing index, the mathematics national examination package 3324 for the senior secondary school students in East Nusa Tenggara Province belonged to the good category. This means that the possibility of the examinees with low ability would be able to answer correctly ten items while the other 30 items which could be

answered correctly without the influence of the guessing factor.

The criteria for the good items for the 3 logistic parameter must meet three requirements, that is, having a good difficulty index with a range between -2 and 2, having discrimination index in the range between 0 and 2, and having good guessing index with a value of less than 0.2. From the result of the analysis, there were 17 items which were categorized as good. The summary of the item analysis result can be seen in Table 5.

Table 5. The result of the Item Characteristics with 3 Logistic Parameter Model

Item	a	b	c	Category
1	1.669	0.32	0.14	Good
2	0.616	0.674	0.001	Good
3	1.17	0.797	0.137	Good
5	1.004	1.436	0.152	Good
7	0.93	-0.012	0.127	Good
8	1.245	1.159	0.193	Good
9	1.588	0.065	0.024	Good
13	1.415	0.482	0.161	Good
14	1.183	0.992	0.166	Good
15	1.751	1.067	0.053	Good
18	0.903	1.026	0.085	Good
19	0.626	1.608	0.165	Good
26	1.044	1.124	0.073	Good
29	1.32	1.811	0.167	Good
32	1.415	0.482	0.133	Good
35	0.981	1.098	0.143	Good
38	0.978	0.646	0.096	Good

The item information function showed the whole power or contribution of the item in measuring the latent trait measured using the test set. The test information function is the sum of the item information functions. The item information function and the test information function are closely related to each other, and, so the test information function will be high when the item information function is also high. A test with high information function will result in the small measurement error so that it provides high

contribution to the information function value in revealing the measurement result.

The item estimation using the 3 logistic parameter model on the mathematics national examination package 3324 resulted in the information function value of 58.0128 with a measurement error of 0.1313. The value was obtained when the students' ability parameter (θ) was 1.6. The measurement error in the mathematics national examination package 3324 for the senior secondary school students in East Nusa Tenggara Province in the academic year of 2015/2016 was 0.1313. The measurement error could happen randomly and systematically. Random measurement error was caused by the physical and mental conditions of the examinees and the very difficult materials. Meanwhile, the systematic error measurement was caused by the measurement instrument, that which is measured and that doing the measurement. The test information function value graph with the parameter of examinees' parameter is presented in Figure 1.

The summary of the analysis result of the mathematics national examination test package 3324 for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province can be seen in Table 9. The summary of the analysis result can be in the form of test item parameter characteristics, the ability characteristics, the information function, and the measurement error.

Table 6. The Item Characteristics Result using the 3 Logistic Parameter Model

Theory	Characteristics
IRT	The mean of the the item difficulty level 2.565
	The mean of the discrimination index 1.543
	The mean of the artificial guessing index 0.135
	The test maximum information function was 58.0128 at the ability of 1.6
	SEM = 0.1313

Cheating Analysis

The analysis of the act of cheating in this study used five methods, that is, the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA) method, and the G2 method. The analysis was done in 22 districts or municipalities in East Nusa Tenggara Province. The findings showed that students in 14 districts were indicated to do cheating. The 14 districts were Kupang Municipality, Kupang district, Timur Tengah Selatan district, Manggarai district, Sumba Barat Daya district, Timur Tengah Utara district, Manggarai Barat district, Manggarai Timur district, Malaka district, Belu district, Rote Ndao district, Sabu Raijua district, Lembaya district, and Flores district.
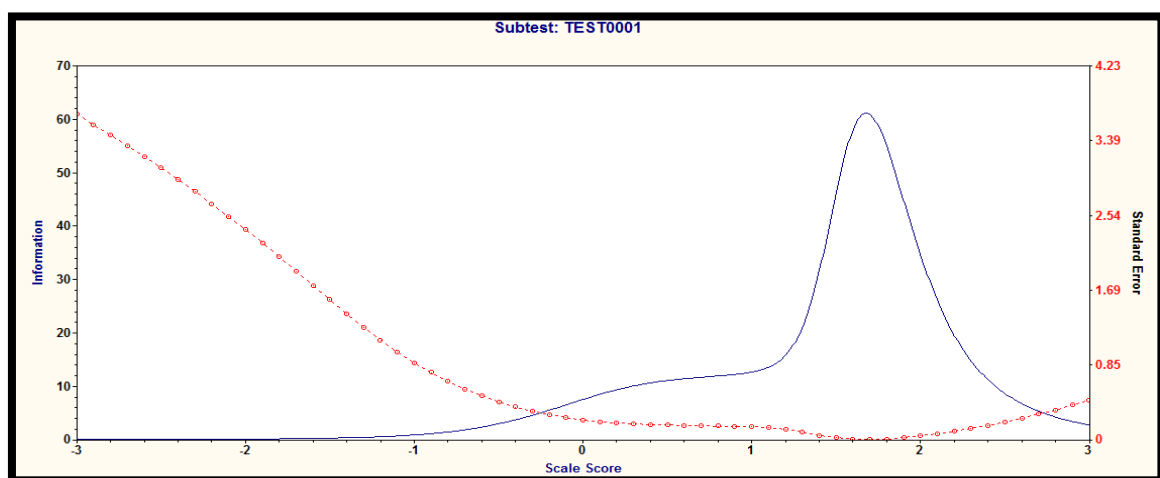

Figure 1. The Information Function Graph, the Ability Scores, and the Standard Error

There was no indication of the act of chearing in 8 districts, that is, Ende district, Sumba Timur district, Sumba Tengah district, Nagekeo district, Sikka district, Sumba Barat district, and Ngada district. The summary of the whole analysis result of the pairs doing cheating in every district/municipality in East Nusa Tenggara Province is presented in Table 7.

Table 7. The Number of Students Doing Cheating in Every District/Municipality in East Nusa Tenggara Province

District/municipality	Number of pairs
Kota Kupang	52
Kabupaten Kupang	63
TTS	67
TTU	4
Belu	8
Malaka	5
Rote Ndao	2
Lembata	1
Sabu Raijua	6
Manggarai	16
Flores Timur	1
Manggarai Barat	7
Manggarai Timur	45
Sumba Barat Daya	17
Jumlah	294

Table 5 presents the frequency of the whole pairs indicated to do cheating.

From the table, it can be seen that the highest number of pairs doing cheating was found in Kupang municipality, Kupang district, and Timur Tengah Selatan district.

In addition, the analysis result of the cheating detection for the whole pairs indicated to have done cheating based on each method, that is, the high category, the medium category, and the low category, is presented in Table 7.

The number of the pairs doing cheating as detected by the G2 method was 183 pairs. This was the highest number of pairs detected. The G2 method detected the cheating based on the number of correct answers and the number of incorrect answers from the pairs suspected to be involved in the act of cheating. In the next order was the Pair 2 method. This method could detect 141 pairs doing cheating. This method did the detection by looking at the number of the items having the same answers on successive items. In the third order was the MESA method which detected 67 pairs detected as doing cheating. This means that the probability of the pairs to choose the wrong answers in the same test items was as many as 67 pairs.

Table 8. The Pairs Indicated to do *Cheating* by Methods

Category	The method detecting *Cheating*				
	B-Index	Pair 1	Pair 2	MESA	G2
High	7	0	57	0	26
Medium	15	6	32	5	76
Total	41	46	52	62	81
Total	63	52	141	67	183

In the fourth order, the method which could detect a great number of pairs was the B-Index method. This method could detect 63 pairs. In detecting pairs doing cheating, this method made a comparison the number of incorrect identical responses to all pairs included in the same interval from the test result. The interval here was based on the multiplication of the number of incorrect responses for 63 pairs of the examinees. Meanwhile, the Pair 1 method was the method which could detect the least number of pairs doing cheating, that is, 52 pairs. This means that the response copying level in the closest pairs was as many as 52 pairs.

The analysis result of the pairs indicated to have done cheating based on the closeness of the seating can be seen from the students' number in one school as presented in the output of the *Integrity*^{TM} software. The analysis result of the pairs indicated to have done cheating can be seen in Table 9.

Table 9. The Analysis Result of *Cheating*
Based on Seating Closeness

Category	*Cheating* detection method				
	B-Index	*Pair 1*	*Pair 2*	*MESA*	*G2*
High	3	1	52	0	29
Medium	4	6	28	5	73
Low	10	41	28	58	74
Total	17	48	108	63	176

Tabel 7 presented that most pairs were indicated doing cheating based on seating closeness. The G2 method detected the highest number of pairs, followed by the Pair 2 method, the MESA method, the Pair 1 method, and B-Index method successively. The high number of pairs was detected by the G2 method because this method assumed that it is possible for each student to have his or her own response alternative when doing a test.

The analysis result also showed the possibility of cheating happening among pairs from different places or different schools in one district. This indicated that it was possible for some parties to distribute the answer keys in different schools located in districts or municipalities in East Nusa Tenggara Province based on benchmark groups. The result of the analysis is presented in Table 10.

The information obtained from Table 13 was the big number of pairs indicated to have done cheating based on the seating closeness. If this is compared to Table 7, the number of pairs indicated to have done cheating based on seating closeness had the highest frequency compared with the benchmark group. This means that the pairs indicated to have done cheating in the administration of the mathematics national examination were dominated by the behaviour of copying others' work based on seating closeness. Although the number of the cheating behaviour based on benchmark group was small, there was an indication of the distribution of answer key to the examinees.

Table 10. The Anlaysis Result of the
Cheating Behaviour based on the
Benchmark Group

Category	*Cheating* detection method				
	B-Index	*Pair 1*	*Pair 2*	*MESA*	*G2*
High	4	0	5	0	0
Medium	11	0	4	1	1
Low	31	4	24	4	7
Total	46	4	33	5	8

The result of the analysis also provided information about which method could detect the cheating behaviour accurately viewed from the standard error. The computation of the standard error in this study used the standard deviation from each method. The method with the smallest standard error was the most accurate method. The result of the computation is presented in Table 9.

Table 11. The Accuracy of the Cheating
Detecting Method

Method	*SD*	*Sample*	*SE*
B-Indeks	0	270888	0
Pair 1	245.261	270888	0.472
Pair 2	557.783	270888	1.072
MESA	33.167	270888	0.064
G2	13.2451	541776	0.018

Table 11 presents the information about the most accurate method in detecting the cheating behaviour. The most accurate method was the B-Index. There was no standard error in this method because there was no standard deviation in this method. The next method was the G2 method because the standard error was only 0.018. The next method was the MESA with the standard error of 0.064. The next method was the Pair 1 method with the standard error of 0.472 and the last is the Pair 2 method with the standard error of 1.072.

As a whole, the cheating detection method analysis is presented in the following figure (Figure 6). In the figure, it can be seen

that the number of pairs detected by the Angoff's B-Index method was 63 pairs, that by the Pair 1 method was 52 pairs, that by the Pair 2 method was 141 pairs, that by the MESA method was 67 pairs, and that by the G2 method was 183 pairs. The pairs here were those sitting close to each other and those pairs who were not at the same place during the administration of the examination but they did cheating were included in the benchmark group to suspect that there was an indicatioin of the distribution of the answer key.
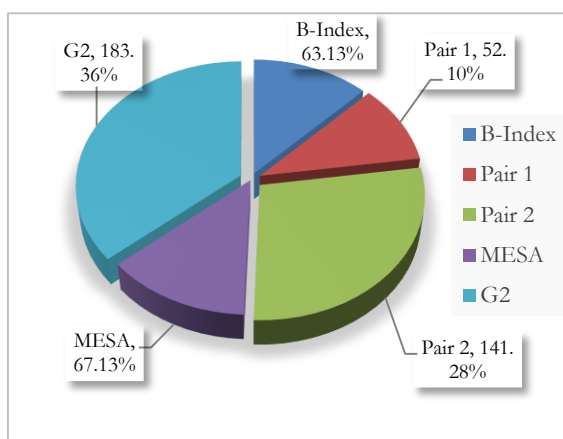

Figure 2.   Figure 2. The whole pairs indicated to have done cheating

In the theory of measurement in educational field, the cheating behaviour is not expected to happen because the cheating behaviour will provide inaccurate information about one's ability. In addition, the cheating behaviour is not fair for other students who do not do cheating. As a result, the students who do not do cheating felt that they were treated unfairly. The cheating behaviour also influence the students' lives in the community. This will result in dishonest generation. Ironically, after doing a crosscheck by looking at the response pattern of a pair of students sitting close to each other, that is, student no 170 and student no 172, it was found that the two students had similar answers from item no 1 to the last item. This also indicated that the proctor did not remind the students not to do cheating in this school.

The essense of the problem of cheating behaviour in the administration of the mathematics national examination for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province was not only because the scores were used as the graduation criteria, but because of the test form which made it possible for the examinees to do cheating. The test form which is sensitive to the cheating behaviour is the multiple-choice test. Many people say that most examinees seem to copy others' work in the examination because it is easy to do this. The examinees can easily work together with other examinees through using symbols found in the answer sheet. Those symbols facilitate other examinees to do cheating because it does not take a long time to get the answer, let alone with the advancement of technology.

Conclusion

Based on the result of the analysis and the discussion on the mathematics national examination package 3324 for the senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province, it can be concluded: (1) the analysis result based on the Item Response Theory showed that out of 40 items analysed using the 3 logistic parameter model, there were only 17 (42%)j items considered as good with a test information function of 58.0128 at the students' ability (θ) of 1.6 and SEM of 0.1313; (2) based on the analysis, there were 63 pairs detected using the Angoff's B-Index method, 52 pairs detected using the Pair 1 method, 141 pairs detected using the Pair 2 method, 67 pairs detected using the MESA method, 183 pairs detected using the G2 method; (3) based on the analysis result, the method which could detect from the highest number of pairs tro the lowest number of pairs doing cheating in the administration of the mathematics national examination in East Nusa Tenggara Province in the academic year of 2015/2016 was the G2 method, the Pair 2 method, the MESA method, the Angoff's B-Index method, and the Pair 1 method successively; (4) based on the computation of the standard error, the most accurate method being able

to detect cheating in the administration of the mathematics national examination in East Nusa Tenggara Province in the academic year of 2015/2016 was the Angoff's B-Index method, the G2 method, the MESA method, the Pair 1 method, and the Pair 2 method successively.

Based on the study, there are some suggestions related to the mathematics national examination test set package 3324 for senior secondary school students in East Nusa Tenggara Province as follows: (1) For the government, it is suggested that the test items developed should be developed so as to meet the good standard so that it can provide accurate information about students' ability. It is suggested that in the future, the test be tried out in remote areas; (2) For schools and teachers, it is suggested that they instill honesty in doing the national examination because cheating behaviour is not an alternative in achieving good scores but through implementing innovative teaching systems, effective teaching methods so that cheating behaviour can be avoided; (3) experts in the educational field are expected to develop accurate detecting method based in good criteria so that the result of the cheating detection can be more accurate.

References

Anderman, E., & Murdock, T. (2007). *Psychology of academic cheating.* Cambridge: Academic Press.

Angoff, H. W. (1972). *The development of statistical indices.* New Jersey, Berkeley: Educational Testing Service Priceton.

Badan Standar Nasional Pendidikan. (2015). *Buku saku tanya jawab pelaksanaan UN tahun 2016.* Jakarta: Badan Standar Nasional Pendidikan.

Baird, J. S. (1980). Current trends in college cheating. *Psychology in the Schools, 17*(4), 515–522. https://doi.org/10.1002/1520-6807(198010)17:4<515::AID-PITS2310170417>3.0.CO;2-3

Bernardi, R. A., Baca, A. V., Landers, K. S., & Witek, M. B. (2008). Methods of cheating and deterrents to classroom cheating: an international study. *Ethics & Behavior, 18*(4), 373–391. https://doi.org/10.1080/10508420701713030

Bernardi, R. A., Banzhoff, C. A., Martino, A. M., & Savasta, K. J. (2012). Challenges to academic integrity: identifying the factors associated with the cheating chain. *Accounting Education, 21*(3), 247–263. https://doi.org/10.1080/09639284.2011.598719

Biro Komunikasi dan Layanan Masyarakat Kementrian Pendidikan dan Kebudayaan. (2016). *Indeks integritas ujian nasional (IIUN) SMA 2016 meningkat.* Jakarta: Kementrian Pendidikan dan Kebudayaan.

Bopp, M., Gleason, P., & Misicka, S. (2001). *Reducing incidents of cheating in Adolescence.* Master's thesis. Saint Xavier University.

Brimble, M., & Clarke, P. S. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *Australian Educational Researcher, 32*, 19–44.

Calabrese, R. L., & Cochran, J. T. (1990). The relationship of alienation to cheating among a sample of American adolescents. *Journal of Research & Development in Education, 23*(2), 65–72.

Cizek, G. J. (2001). An overview of issues concerning cheating on largescale tests. In *The Annual Meeting of the National Council on Measurement in Education.* Seattle, WA.

Davis, S. F., Drinan, P. F., & Gallant, T. B. (2009). *Cheating in school; what we know and what we can do.* United Kingdom: Wiley-Blackwell.

Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: prevalence, determinants, techniques, and punishments. *Teaching of Psychology, 19*(1), 16–20. https://doi.org/10.1207/s15328023to

p1901_3

DeMars, C. (2010). *Item response theory; undestanding statistics measrument.* New York: Oxford University Press.

Ferry, B. R., Tidman, M. T., & Wats, M. T. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics, 2*(4), 235–256.

Hanson, A. B., Haris, J. D., & Brenan, L. R. (1987). *A comparison of several statistical methods for examining allegations of copying.* Iowa: American College Testing.

Hughes, J. M. C., & McCabe, D. L. (2006). Academic misconduct within higher education in Canada. *Canadian Journal of Higher Education, 36*(2), 1–21.

Jensen, A. L., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybodydoes it: Academic dishonesty among high school and college students. *ContemporaryEducational Psychology, 27,* 209–228.

Lin, C. S., & Wen, L. M. (2007). Academic dishonesty in higher education: A nationwidestudy in Taiwan. *Higher Education, 54,* 85–97.

Manoppo, Y., & Mardapi, D. (2014). Analisis metode cheating pada tes berskala besar. *Jurnal Penelitian Dan Evaluasi Pendidikan, 18*(1). Retrieved from https://journal.uny.ac.id/index.php/jpep/article/view/2128

Mardapi, D. (2000). Evaluasi penyelenggaraan ebtanas. *Jurnal Kependidikan, XXX*(2).

Michaels, J. W., & Miethe, T. D. (1989). Applying theories of deviance to academic cheating. *Social Science Quarterly, 70*(4).

Mulyati, B., & Kartowagiran, B. (2013). Analisis hasil ujian nasional mata pelajaran ekonomi SMA di kota Serang. *Jurnal Evaluasi Pendidikan, 1*(1). Retrieved from http://journal.student.uny.ac.id/ojs/index.php/jep/article/view/45

Naghdipour, B., & Emeagwali, O. L. (2013). Students' justifications for academic dishonesty: call for action. *Procedia - Social and Behavioral Sciences, 83,* 261–265. https://doi.org/10.1016/j.sbspro.2013.06.051

Newstead, S. E., Franklyn-Stokes, A., & Armstead, P. (1996). Individual differences in student cheating. *Journal of Educational Psychology, 88*(2), 229–241. https://doi.org/10.1037/0022-0663.88.2.229

Presiden Republik Indonesia. Peraturan Pemerintah RI Nomor 13, Tahun 2015, tentang Perubahan Kedua atas Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan (2015).

Rakovski, C. C., & Levy, E. S. (2007). Academic dishonesty: Perceptions of business students. *College Student Journal, 41,* 466–481.

Stogner, J. M., Miller, B. L., & Marcum, C. D. (2013). Learning to E-Cheat: a criminological test of internet facilitated academic cheating. *Journal of Criminal Justice Education, 24*(2), 175–199. https://doi.org/10.1080/10511253.2012.693516

Vandehey, M. A., Diekhoff, G. M., & LaBeff, E. E. (2007). College cheating: A twenty-yearfollow up and the addition of an honor code. *Journal of College Student Development, 48,* 468–480.

Whitley, B. E., Nelson, A. B., & Jones, C. J. (1999). Gender differences in cheating attitudes and classroom cheating behavior: A meta-analysis. *Sex Roles, 41*(9), 657– 680.

Zastrow, C. (1970). Cheating among college graduate students. *The Journal of Educational Research, 64*(4), 157–160. Retrieved from http://www.jstor.org/stable/27536093