MEASUREMENT ERROR ESTIMATION OF CUT SCORE OF ANGOFF METHOD BY BOOTSRATP METHOD

Sebastianus Widanarto Prijowuntato

Graduate School.

Yogyakarta State University, 2014

Abstract

The purpose of the study was to determine cut score of National Examination in Accounting Subject of Package 2 and to estimate the measurement error in the Angoff The sample were 12 SMK's in DIY purposively selected. Teachers who follow a Focus Group Discussion were 9 people consisting of 7 women and 2 men. Data obtained by documentation of a national competency exam answers Accounting of package 2 in the academic year 2011/2012. The technical analysis was divided into three stages. The first stage was the preparation. The activities at this stage included the preparation of data. The second stage was FGD, which was carried out in two rounds. The FGD participants in the first round were given a training to determine the cut score using the Angoff method. In the second round, the participants specified a cut score but they had not been given a training anymore. At the third stage, the participants estimated error measurement by using the Bootstrap method. The steps used in the Bootstrap method included the determination of the data, resampling the sample (x^*), calculating the standard error of estimation of Bootstrap. The results showed that 1) the cut scores for the Angoff was 68.22. 2) The error estimation of Angoff's cut score was 1.58 in 200 times Bootstrap.

Keywords: *Cut Score, Error Measurement Estimation, Bootstrap*

Introduction

The Indonesian government concern about the quality of education. This is evident in the policies issued by the government, especially in the allocation of education funding in the *APBN*, the Law of National Education System, and educational standards such as content standard, competency standard, facilities and infrastructure standard, management standard, process standard, financing standard, educators and education personnels standard, and assessment standard.

Recently, the government sets the graduation policy at every level of education. At this time, the student pass on the educational level when they reach 5.5 on the national exam. If the student does not reach 5.5, the student does not pass and have to take the exam for elementary equivalence package A, package B for junior high, and package C to high school.

Cut scores determination is not easy (Nudell, 2008). For example, before determining the cut scores, standard setters have to 1) agree about minimum competencies definition, that must be achieved by students, 2) determine the number of panelists who are involved in the standard setting. 3) determine the number of rounds, 3) determine cut score if there are

differences the cut score in each round. These variables can lead to variability in the resulting cut scores. Therefore, standard setters need to pay attention to the existing standard error on cut scores.

A number of studies is held to find cut score (eg. Alsmadi, 2007; Skagg, Hein, & Awuor, 2007; Natalina, 2010; Koffler, 1980; MacNamm & Stanley, 2006). However, a few research considered standard error in determining cut score. That impact in the cut score precision as cut of point student competencies.

Angoff method broadly used in standard setting (MacCann & Stanley, 2004). Researcher use Angoff method to find cut score and compare with another standard setting method (Yin and Schulz, 2005; Skaggs, Hein, and Awuor, 2007; Alsmadi, 2007; Brennan & Lockwood, 1980). Using Angoff method, panelists ask to estimate minimal competency probability of students who can answer items correctly.

Panelists rarely estimate of minimal competency probability preciously. Nichols, Twing & Mueller (2010) suggested that a problem in measuring process in social science is no indicator that measure the attribute directly. The accuration in estimation of minimal competency probability can be approached by Bootstrap method. The Bootstrap method is used with resample with replacement repeatedly.

Some literature suggests that resulting resample with replacement repeatedly are close to the normal distribution. By using the Bootstrap method, standard setters can calculate the standard errors. The Bootstrap method is widely applied to the statistics science for mengestimasi errors in small populations or population numbers were not known.

According to Efron, & Tibshirani (p. 45, 1993), resampel with replacement for estimating of standard error is done at least 200 times. In this study, resample of the Bootstrap would have done as many as 200, 300, 500, and 1000 times. It is meant to see the differences among Bootstrap results. Standard error is calculated based on the results of the resample.. Results calculation of the standard error will be close to the original sampel.

The Bootstrap steps for estimating of standard error are as follows (Ajmani, 2009, p. 264).

1. Determine reseach population, in this case is cutscore that generated through Focus Group Discussion
2. Take sample (x) from existing population.

3. Take resample (x^*) with replacement as many as n = 7 from the sample (x). Resample with replacement is hold as many as 200, 300, 500, and 1.000 times.

4. Calculate the estimation of Bootstrap standard error by formula (Efron & Tibshirani, 1993. p.47).

$$se_B = \left\{ \sum_{b=1}^{B} [\hat{p}^*(b) - \hat{p}^*(.)]^2 / (B-1) \right\}^{1/2}$$

where:

se_B = Bootstrap standard error
$\hat{p}$ = parameter of population
$\hat{p}^*(.) = \sum_{b=1}^{B} \hat{p}^*(b) / B$ (mean of $\hat{p}_1 ... \hat{p}_B$)
B = number of resample

The process of Bootstrap in this research is drawn in Figure 1. In Figure 1, sample $(x_1, x_2, ... , x_n)$ that is used in Bootstrap is derived from real population (p). Resampling with replacement is taken from the original sample (n). The star notation indicates that (x*) is not the actual data set x, but rather a randomized, or resampled, version of x. Bootstrap is done as many B times. The next step is calculate the statistics of each sample Bootstrap standard error.

Based on the background, the proposed research question is how large standard error on the Angoff method? This research aims to estimate the measurement error in Angoff method by Bootstrap on accounting expertise field of vocational school in Yogyakarta in the academic year of 2011/2012.
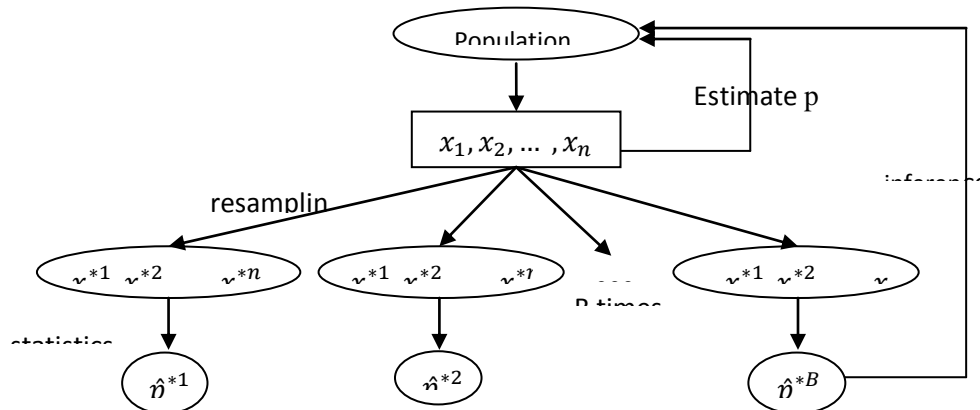
Gambar 1.

This research contribute to estimate standard error that generated using Bootstrap. The other research estimate standard error by using Central Limit Theorm (MacCann, & Stanley, 2004), and generalizability Theory (Yin, & Sconing, 2008).

Reseach Method

The reseach data were students' response to the National Examination in Accounting Subject of Package 2 of accounting expertise field, study program financial of vocational school, field expertise business and management of vacational school in Yogyakarta in the academic year of 2011/2012 which are schools' data with 338 students.

The population of the reseach were cut score that are generated using Angoff method on vocational school in Yogyakarta. The sample of the research were cut score of nine vocational school that have accounting expertise field in Yogyakarta.

The reseach uses instrument of Practice Examination in Accounting Package 2 of vocational school created by National Education Standards Agency (BNSP). The question of practice examination consist of three parts namely managing journal, ledgers, and accounting cycle. In the managing journal there are 26 transaction that must be done by examinee. After the examinees answer the question in the first section, the examinee are asked to post a journal to the ledgers. In third section, examinees are asked to complete accounting cycle by made the Bank reconciliation journal, the adjustments, income statements, statements of equity, the balance sheet and cash flow, closing journal, and balance sheet after closing journal.

Technique of Data Analysis

The technique of data analysis was divided into three stages. The first stage was the preparation. The activities at this stage included the preparation of data. In this stage, researcher prepared data that will be used in standard setting meeting. The data included test items, examinee response of national examination, examinees score, cut score worksheet, material of cut score training.

The second stage was Focus Group Discussion (FGD), which was carried out in two rounds. The 12 teacher of vocational school has invited to standard setting meeting. Teacher who engaged in the FGD were nine teachers, consisting of seven women and two men. The FGD participants in the first round were given a training to determine the cut score using the Yes/No Angoff method. Procedur of cut score determination are as follows. 1) asked to teachers to review test items, 2) asked to the teacher to estimate examinee who answered test

items correctly. If amount of examinees answered test item correctly than they got score 1. If amount of examinees could not answer test item correctly, than they got score 0. 3) made avarage of test score. In the second round, the participants specified a cut score but they had not been given a training anymore.

At the third stage, the researcher estimated error measurement by using the Bootstrap method. The steps used in the Bootstrap method included the determination of the population, sampling the data, resampling the sample (x^*), calculating the standard error of estimation of Bootstrap. The formula to find standard error using Bootstrap as follow (Efron, & Tibshirani, 1993. p.47).

$$se_B = \left\{ \sum_{b=1}^{B} [\hat{p}^*(b) - \hat{p}^*(.)]^2 / (B-1) \right\}^{1/2}$$

In this stage, researcher made resample the cut score as much as 200, 300, 500, and 1.000 times using R program serie i386 3.0.0. The syntax used to resample are as follow.

```
> getwd()
> data=read.csv("angoff.csv", header=T)
> data=angoff.vector
> data=data$Cutscore
> data
> sink("angoff.txt")
> for (i in 1:200){
+ x=b.stat(data,200,mean)
+ print(c(i,mean(x$stats), x$std.err))
+ }

> sink()
```

Researcher also made a file to compute standard error. The syntax of the file are as follow.

```
b.stat <- function(data, num, stat) {
resamples <- lapply(1:num, function(i) sample(data, replace=T))
r.stat <- sapply(resamples, stat)
std.err <- sqrt(var(r.stat))
list(std.err=std.err, resamples=resamples, stats=r.stat)
}
```

Research Finding and Discussion

Data used in data analysis are students' response toward the National Examination in Accounting Subject of Package 2 of accounting expertise field, of vocational school in Yogyakarta in the academic year of 2011/2012. There are constrain in the data collection, 1) principle of school has no saved the bundle of national examination anymore. 2) some school take the package 1 and 3 for national examination. 3) a few school are fishing out. The 12 schools that are became as data source represented: state and private school, regency in Yogyakarta, and highest, middle, and lowest catagorical school. Tabel 1 indicate the sample school.

The teachers who attended in FGD are one teacher from SMK PGRI 1 Sentolo, one teacher from SMK Tujuh Belas Bantul, one teacher from SMK Budhi Dharma Piyungan, one teacher from SMK Negeri 7 Yogyakarta, one teacher from SMK Muhammadiyah Tepus, one teacher from SMK Sanjaya Gunung Kidul, one teacher from SMK Sanjaya Pakem, one teacher from SMK YPKK 3 Sleman, dan one teacher from SMK "17" Seyegan. The nine teachers are in accordance with the specified conditions that are have been teaching for a minimum of 5 years, teaching in grade 12 and graduate of the Economic/Accounting Studi Program.

Tabel 1. School List

No	School Name	Catagory	No	School Name	Catagory
1.	SMK Ma'arif 1 Temon	Low	7.	SMK Muhammadiyah Tepus	High
2.	SMK PGRI 1 Sentolo	High	8.	SMK Sanjaya Gunung Kidul	High
3.	SMK Tujuh Belas Bantul	Low	9.	SMK Sanjaya Pakem	Middle
4.	SMK Budhi Dharma Piyungan Bantul	Middle	10	SMK Muhammadiyah Berbah	Middle
5.	SMK Negeri 1 Bantul	High	11.	SMK "17" 1 Seyegan	Low
6.	SMK Negeri 7 Yogyakarta	High	12	SMK YPKK 3 Sleman	Low

Stage I

Standard setting activities begins with preparing devices used in standard setting. This preparation includes 1) to correct a student's response of vocational school in the dichotomous form. The corrector are student of accounting education at semester 9 of Sanata Dharma University. Before correcting examinee response, the students of accounting education are trained in advance to identify perceptions. Assessment rubrics for dichotomous scoring also socialized to students. The materials distributed to standard setting participants consists of a description of standard setting, standard setting rounddown, panelists answer

sheet form, the standard of competence of graduates (SKL), vocational practice examination questions, and assessment rubrics.

Stage II

FGD conducted with 12 panelists from each sample school. Before determining cut score, panelist was given an explanation about understanding and purposing the standard setting, the instrument used, and data analysis using Angoff method. The participants also were given hard copy material. After an explanation of the standard setting material, participants practice standard setting under the guidance of researcher. The cut score are simulated using Excell program. After that, participants determine the cut score using Yes/No Angoff method.

Determination of cut score using Angoff method is condusted by gave a score in each descriptor (journal, ledgers, and accounting cycle). If examinees are able to answer the question then the examinee is given a score of 1, whereas if the examinee are not able to answer the question then examinee is given a score of 0. In the first round, cut score generated by the panelists is 68.44 and in the second round, the cut score generated by the panelists is 68,22. The cut score is shown in Tabel 2.

Tabel 2 Cut Score of Yes/No Angoff Method

	Round 1	Round 2
Panelist 1	72.00	72.00
Panelist 2	68.00	67.00
Panelist 3	72.00	73.00
Panelist 4	73.00	72.00
Panelist 5	70.00	70.00
Panelist 6	64.00	60.00
Panelist 7	54.00	60.00
Panelist 8	73.00	73.00
Panelist 9	70.00	67.00
Mean	68.44	68.22

Based on Tabel 2, it appears that the cut score generated by the panelists include high category. Panelist 6 and 7 gave the lowest cut score, i.e 60.00. while the panelist 3 and 8 gave the highest cut score, i.e 73.00. The average overall cut score was 68,22. This showed that examinees are considered to be competen if they reached score minimum 68.22 or more.

If the standard setter use the cut score 68.22 to determine pass/fail, there are 111 of 338 examinees who graduated from the national examination. Thus, there are 227 examinees who did not pass. It appears that the cut score generated in standard setting is high. It means that examinee who graduated from vocational school have a competency to manage accounting process. This understandable that the accounting field is one part of the company that holds an important role. The existing errors in one of accounting process would result in the next process and it made a company loss. Therefore, the business expected graduate of vocational school expert in their field.

Results of discussion with the teachers revealed that the time provided to complete the accounting practice examination was not proportional. The time given to complete managing journal was shorter than others (ledgers and accounting cycle). While, questions that given in managing journal was more difficult than others.

Determination of cut off score that high needs to be balanced with a good learning process, good facilities and infrastucture, and the involvement of various party in learning. in practice, sometimes, there is a disharmony between the company and the education. For example, if student do field work practice, many company divert the accounting students in other field, such as secretariat, sales force, etc. Many company assumed that finance is company confidential and anyone had not to know.

The school's education facilities and infrastructure need to be improved. Based on survey conducted by researcher, many school have no adequate facilities and infrastructure for learning well. Many school had limited class, consequently, it had no accounting laboratory. In addition, the tools of learning in vocational schools was minimal.

The lack of existing facilities and infrastructure will result in less then optimal learning and its impact on the poor quality of vocational graduates. Therefore, in addition to the teacher demanded more creative in the learning of accounting to their students. Teachers also need to introduce accounting practice and bring it in the classroom. Therefore, the teachers' creativity needs to be improved by providing a good education and training.

The government should concern about the conditions above and should increase education quality. The government should pay more attention to the quality of the schools that is in standard below category. The gap between the business and education can be minimized by the existence of coordination between the ministries concerned. The government also provide control for schools, especially vocational school either administration, teachers, the learning process through the educational distric agency.

Stage III

Cut score (x) that generated on stage II is taken resample with replacement (x*). The resampling is conducted using the R i386 3.0.0 program. Tabel 3 indicated estimation of standard error. Tabel 3 showed that second round have less standard error compared by first round.

Tabel 3 Estimasi Kesalahan Standar pada Metode *Yes/No Angoff's*

Bootstrap	Putaran 1	Putaran 2
200	1.90	1.58
300	1.91	1.63
500	1.90	1.69
1000	1.86	1.63

Estimation of standard errors obtained using Bootstrap is relatively small. Several factors are thought to be a large effect on the small standard errors is the ability of panelists in this case teacher of vocational school. Teachers who understand the students' ability are easily predict the students ability. It can be held by discussion frequently about subject or other knowledge.

The other factor thought to affect the estimating of the standard error is the number of sample. It influence in variation data. The more sample that are made on the Bootstarp resample the data will be more variation. Similarly, the less sample that is resampled, the data obtained did not have many variation.

In addition, the cut score made by teachers affected by definition and training of standard setting (Giraud, Impara, & Plake, 2005). Therefore, in conducting standard setting, the teachers have to practice and sufficient training. Understanding in standard setting well have an impact on the determination of cut score.

Conclusion and Suggestion

The finding suggest that cut score for National Examination in Accounting Subject of Package 2 was 68,22. Further, the cut score was higher than Indonesia government determined. Bootstrap could be used for estimate the cut score error. The estimation of cut score error was various depended on amount boostrap.

There is a need for furhter studies of the estimation of cut score error. For example comparing error estimation of Angoff cut score and other cut score method by bootstrap method. Futher studies might comparing bootstrap error estimation and other method.

References

Alsmadi, A. Abdalla, 2007. A Comparative Study of Two Standard-Setting. *Social Behavior and Personality*. 38 (4). pp. 479 – 486.

Brennan, R. L., & Lockwood, R. E., 1980. A Comparison of Nedelsky and Angoff Cutting Score Procedures Using Generalizability Theory. *Applied Psychological Measurement*. 4 (2). Spring. pp. 219 – 240.

Efron, B., & Tibshirani, Robert J., (1993). *An Introduction to the Bootstrap*. Chapman & Hall. Inc.. New York.

Giraud, G., Impara, James C., & Plake, Barbara S., Teachers' Conception of the Target Examinee in Angoff Standard Setting, *Applied Measurement Education*, 18 (3), pp. 223 - 232

Koffler, Stephen L.. 1980. A Comparison Approach for Setting Proficiency Standards. *Journal of Educational Measurement*. 17 (3). Autum. pp. 167 - 178

Macnamm, Robert G.. & Stanley, G.. 2004. Estimating The Standard Error of Judging in Modified-Angoff Standard Setting Procedure. *Practical Assessment, Research, & Evaluation*. 9 (5). Retrieved November 14, 2013 from http://PAREonline.net/getvn.asp?v=9&n=5

Macnamm, Robert G.. & Stanley, G.. 2006. The Use of Rasch Modeling to Improve Standard Setting. *Practical Assessment, Research & Evaluation*. 11 (2). January. pp. 1 – 17.

Natalina Premastuti B.. 2010. Komparasi Standard Setting Metode Group Contrast dan Bookmark pada Mata Pelajaran Akuntansi. *Penelitian dan Evaluasi Pendidikan*. 14 (2). pp. 226 – 245.

Nichols, P., Twing, J., & Mueller, C. D. (2010). Standard-Setting Methods as Measurement Process. *Educational Measurement: Issues and Practice,* Vol. 29, No. 1, Spring, pp. 14 – 24.

Skagg, G. H., Serge F.. & Awuor, R. 2007. Setting Passing Score on Passage-Based Test: A Comparison of Traditional and Single-Passage Bookmark Method. *Applied Measurement In Education*. 20 (4). pp. 405 – 426.

Yin, P., & Schulz, E. M. (2005). A Comparison of Cutscores and Cutscore Variability from Angoff-Based and Bookmark-Based Procedures in Standard Setting. *Paper presented at the Annual Meeting of the National Council on Measurement in Education*, Montreal, Canada.

Yin, P., & Sconing, J. 2008. Estimating Standard Errors of Cut Scores for Item Rating and Mapmark Procedure: A Generalizability Theory Approach. *Educational Psychological Measurement*. 68 (1). February. pp. 25 – 41.