Kirkpatrick Evaluation On Education And Training Of School Supervisor based On Best Practice, Hots And Adult Education In Indonesia

Renny Yunus[1], Siti Patimah[2], Nirva Diana[3], Agus Pahrudin[4] , Koderi[5]

1 Student of Post Graduate Doctoral Program UIN Raden Intan Lampung, Indonesia
2,3,4,5,Lecturers of UIN Raden Intan Lampung, Indonesia
* Corresponding Author:
Email: artemiloveyousufi@gmail.com

Abstract.
Education quality improvement is a joint responsibility that involves various groups, from structural aspects starting from the central level (ministry) to the education unit. Law Number 20 of 2003 on National Education System Article 39 Paragraph 1 states that education personnel have the duty to administer, manage, develop, supervise, and provide technical services to support the education process in educational units. To achieve the National Medium Term Development Plan - Strategic Plan of Directorate General of Teachers and Education Personnel 2015-2019, the Ministry of Education and Culture has set the Average Target of Knowledge and Skills Competency. This purpose of this study was to evalaute the education and training of school supervisor by using Kirkpatrick evaluation. The results of this study recommended that the supervision can be improved by using education and training based on HOTS, Best Practice and Adult Education.

Keywords *: Education and training, kirkpatrick evaluation, school supervisor*

I. INTRODUCTION

Improvement of education quality is a joint responsibility that involves various groups, from structural aspects starting from the central level (ministry) to the education unit [1].Law Number 20 of 2003 on National Education System Article 39 Paragraph 1 states that education personnel have the duty to administer, manage, develop, supervise, and provide technical services to support the education process in educational units. Based on the pre-survey, despite attending education and training, the performance of the School Supervisor was highlighted because it failed to improve the quality of the institution and even decreased in the past 2 years. This is inversely proportional to the general objectives and special objectives of the education and training program of school supervisors in Indonesia according to the Minister of Education Regulation Number 12 of 2007 which has the main task in terms of Academic supervision, Managerial supervision, Educational evaluation, Sustainable Research and development. Because it has such a big role and function, various School Supervisor development efforts have been carried out by the Subdirectorate of Career Development and Performance Assessment, Directorate of Development, Directorate General of Teachers and Education Personnel, Ministry of Education and Culture. To achieve the National Medium Term Development Plan - Strategic Plan of Directorate General of Teachers and Education Personnel 2015-2019, the Ministry of Education and Culture has set the Average Target of Knowledge and Skills Competency of Education Personnel in the Strategic Plan of Directorate General of Teachers and Education Personnel 2015-2019 as in Table 1.2.

Table 1.2.
Strategic Plan of Directorate General of Teachers and Education Personnel 2015-2019
Average Target of Knowledge and Skills Competency of Education Personnel

ACTIVITY AND PERFORMANCE INDICATORS	***BASELINE*** **2014**	**PERFORMANCE TARGET**				
		2015	**2016**	**2017**	**2018**	**2019**

IKK 6.1. Average Competency Scores of Knowledge and Skills of Education personnel in Primary and Secondary Education	4.7	5.6	6.6	7.1	7.6	8.0
IKK 6.2. Number of Education personnel who have improved Performance Index and Attitude Quality		50.87	53.601	107.201	160.801	234.401
IKK 6.3. Number of supervisors of directorate general of primary and secondary education who have graduated masters		400	400	400	400	400

Source : Strategic Plan of Ministry of Education and Culture2015-2019

The process of Education and Training for School Supervisors has been carried out, so that the main tasks of School Supervisor are conveyed and lead to carrying out academic and managerial supervision tasks in the education unit which includes preparation of supervision programs, implementation of mentoring, monitoring of implementation, assessment, guidance and professional training for teachers , evaluation of the supervision program implementation, and the implementation of supervision tasks in special areas. This is in line withSurah Al-Hashr verse18 below

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُواْ ٱتَّقُواْ ٱللَّهَ وَلۡتَنظُرۡ نَفۡسٞ مَّا قَدَّمَتۡ لِغَدٖۖ وَٱتَّقُواْ ٱللَّهَۚ إِنَّ ٱللَّهَ خَبِيرُۢ بِمَا تَعۡمَلُونَ

O you who have believed, fear Allah . And let every soul look to what it has put forth for tomorrow - and fear Allah. Indeed, Allah is Acquainted with what you do [2].

The above verse emphasizes that the supervision program in schools should be directed to support best practice. This is in line with a study by Kotirde& Yunos which realized thatpublic secondary schoolshad supervision problems [3]. Evaluation is an important process that functions to see the process of implementation and success, one of which is the Kirkpatrick Evaluation on education and training for supervisors based on Best Practice, HOTS and adult education in Indonesia. **Kirkpatrick Evaluation can represent every stage of education and training that has been carried out specifically and is able to provide maximum output** results with several items namely:

1. *Reaction Evaluation*
2. *Learning Evaluation*
3. *Behavior Evaluation*
4. *Result Evaluation*

This study used a cross sectional survey with questionnaires at selected secondary schools in Indonesia [4]. Kirkpatrick Evaluation on the implementation of education and training will assist the government in determining professional development programs on the existence of school supervisors in Indonesia. Supervision of principals and teachers involves stimulation of growth and development of teachers and education personnel, selection and revision of educational goals, teaching materials, teaching methods and evaluation of education and training. This can be known through the process of coaching and evaluative.

Therefore, the Ministry of Education and Culture through the Directorate General of Teachers and Education Personnel has prepared a reinforcement pattern with modules and curriculum implemented as a result of the Kirkpatrick Evaluation. At Level 1, the evaluation was carried out through a questionnaire on the components that influence the quality of the implementation of education and training, the quality of instructors and barriers experienced by participants, and suggestions for improvement of subsequent

education and training to then be analyzed qualitatively and descriptively. The evaluation was carried out on the following components namely, learning material, duration, facilities, media, methods used, quality of instructor [5].

At Level 2, the Pre Test and Post Test results were compared in the form of written tests. At this level, the participants of education and training were asked to bring the same education and training material, namely specialization material in the field of participant science which was made into methods and strategies in carrying out their duties as school supervisors. At this level of evaluation, the quantitative data were analyzed by summarizing the results of participant tests in the pre-test and post-test with descriptive and inferential statistical techniques [6]. The evaluation indicators at this level namely:

1. Increasing mastery
2. The effectiveness of learning is determined by increasing mastery of the results of education and training before and after training activities.

Evaluation at this level was performed after completing education and training with a questionnaire containing questions about the use of education and training material that has been followed in carrying out the task [7].

II. METHODS

1. Type and Nature of Research
   a. Type of Research
      This study used mix-type research, which is a combination of field research and library research.
   b. Nature of Research
      This study used descriptive analysis method
2. Source.
   This study used primary data and secondary data.
3. Informant
   In qualitative research, the sample is often referred to as an informant that is the person who is the source of information. The subjects who became informants in this study were academics in UIN Raden Intang Lampung. The determination of informants used a purposive sampling technique.
4. Data Collection Technique
   This study used two methods in collecting data, namely field research and library research. In the field research method, interviews were conducted with informants on the subject matter of the study.
   Library research is a method to collect data by tracking and searching library materials. In this study, the library research was conducted by reading, analyzing, and studying various library materials related to the problem being studied

III. RESULT AND DISCUSSION

Data Sources

This study used simple random sampling from populations without regard to strata in the population. The population was the education supervisors in Indonesia of 16,000 people with a sample of 99 people and rounded up to 100 people as research respondents.

Kirkpatrick Analysis Level 1 Reaction

Instrument Tabulation of Informant

In the informant instrument, there were 6 indicators with feedback from 100 informants. The indicators namely, personality (Kp), managerial supervision ability (KSM), academic supervision ability (KSA), clinical supervision ability (KSK), educational evaluation (EvP), research and development (PP). The scale can be seen as follows, '5' for strongly agree; ‘4’ for ‘agree,‘ 3’ for ‘less agree’; '2' for 'disagree', and '1' for 'strongly disagree'. From the overall graph above it can be concluded that the 6 indicators were assessed by

54 to 110 people who tended to agree with the provisions on the survey form. Based on the parametric test, the data was normally distributed and linear.

Based on the boxplot diagram above, it can be seen that the indicators that obtained approvement namely personality (Kp), academic supervision (KSA), educational evaluation (EvP), clinical supervision ability (KSK), managerial supervision ability (KSM), and research and development (PP).

Kirkpatrick Evaluation Level 2 LEARNING

Instrument Tabulation of Participant

Modul Quality

In the histogram above,majority of participants or 45-57 participants strongly agreed with modul quality. Meanwhile, 50 participants agreed, and under 5 participants less agreed. Based on the normality test, the data was considered normal.

In the boxplot diagram above,the score was between 3 to 5 on 1st, 3rd, 4th, and 7th indicators of modul quality.This shows that the module quality experienced relatively unstable increases and decreases.

Facilitiesof Education and Training

In the histogram of education and trainingfacilities, the score was between 2 to 5. 35-55 participants strongly agreed, 45 to 55 participants agreed, 10 participants less agreed, and under 5 participants disagreed with the facilities of education and training. Based on the normality test, the data was considered normal. This shows that education and training facilities need to be developed.

In the boxplot diagram above,facilitiesof education and traininghad a score of 2 on 5th, 6th, and 7th indicators, but there was an increase on 8th and 9th indicators, which can be shown that there were no participants who disagreed.

Instructor Evaluation

In the histogram of instructor evaluation, the score was between 3 to 5. The highest score of strongly agreed was in the 10th indicator of instructor evaluation. The 'less agree' response had less than 10 participants in the 9th indicator of instructor evaluation. This shows a fairly rapid increase from 9th to 10th indicators of instructor evaluation. Based on the normality test, the data was considered normal.

Implementation Evaluation

In the implementation of evaluation histogram, the results showed a score of 1 or strongly disagree and 2 or disagree. The highest responses of strongly disagree was in the 9th indicator of implementation evaluation which reached nearly 100 participants. The 'less agree' response had less than 5 participants. It showed that implementation evaluation obtained a lot of criticism. Therefore implementation evaluation needs to be taken seriously to achieve the desired work targets.

In the boxplot diagram above, it can be seen that the implementation evaluation had a score with a scale of 1 and 2. The indicators that obtained strongly disagree responses from the participant were 2nd, 3rd, 7th and 8th indicators. While the 1st, 4th, 5th, 6th, and 9th indicators of the implementation evaluation had a disagree response. This shows that the implementation evaluation needs to be improved because the results show a relatively negative responses of the implementation evaluation.

Instrument Tabulation ofInformant

Personality

In the histogram of personality, the results of the questionnaire which showed strongly agree response was in the 4th indicator of personality that reached nearly 90 participants. For less agree responses and agree responses were mostly in the 1st indicator of personality. This shows a good increase from the 1st to 4th indicators of personality.

In the boxplot diagram above, it can be seen that the assessment of personality had a score of 3 to 5. The results of the questionnaire that received a score of 3 to 5 were 1st, 2nd, and 3rd indicators of personality. While the results of the questionnaire that obatined a score of 4 to 5 from the participant was the 4th indicator of personality.

Managerial supervision ability

In the histogram of managerial supervision abilitiy, the score was between 2 and 5. The highest responses of strongly agree was in the 2nd indicator of managerial supervision. The 'less agree' response which had less

than 5 participants was in the 3rd indicator of Managerial Supervision Ability. Based on the normality test, the data was considered normal.

Based on the boxplot in the figure above, it can be seen that the assessment on managerial supervision ability had a score between 2 to 5 in the 3rd indicator. In addition to the 3rd indicator, the 1st, 2nd, and 4th indicators of managerial supervision ability had a score between 3 to 5.

Academic supervision ability

In the histogram of academic supervision ability, the score was between 2 and 5. The highest responses of strongly agree was in the 2nd indicator of academic supervision with 70 participants. While for the disagree response was in the 2nd indicator of academic supervision which had less than 5 participants. This shows that scores that have striking results were found on the 3rd indicator which had the best score as well as the worst compared to other indicators of academic supervison ability. Based on the normality test, the data was considered normal. Based on the box plot in the figure above, it can be seen that besides the 2nd indicator of academic supervision ability, other indicators of academic supervision ability had a score of 3 to 5. All indicators of academic supervision ability were in fairly good category.

Clinical supervision ability

In the histogram of clinical supervision ability, the score was between 3 and 5. The highest response of strongly agree was in the 4th indicator of clinical supervision ability with 70 participants. The highest responses of disagree response was in the 2nd indicator of clinical supervision ability. Based on the normality test, the data was considered normal.

Based on the box plot in the figure above, it can be seen that the assessment on clinical supervision ability had a score with a scale of 3 to 5. All indicators of clinical supervision ability had a stable score and no significant difference.

Educational evaluation

In the histogram diagram and the pie chart of educational evaluation, the score was between 3 and 5. The highest responses of strongly agree response was from 65 participants. The second highest responses was in agree response with 30 participants. While for the disagree response had the lowest participants of under 10 people. Based on the normality test, the data was considered normal.

Research and Development

In the histogram of research and development, the score was between 2 to 5. The highest response of strongly agree was in the 1st indicator of research and development with 50 participants. The highest responses of agree was in the 2nd indicator of research and development. The quality of the modules or the 3rd indicators of research and development had disagree responses of under 5 participants. Based on the normality test, the data was considered normal.

Based on the boxplot diagram in the figure above, it can be seen that the assessment on research and development had a score with a scale of 2 to 5. The only indicator with disagree response was the 3rd indicator of research and development. As a whole, the score of research and development had a decrease.

Kirkpatrick Evaluation Level 3 Behavior

Behaviour – Assessment of Training Indicators

Behavior can be divided into 4 indicators. The first indicator was represented by Tosca color which is a score of 20 training indicators which were relatively stable between 70 to 90. The second indicator was Andragogy with a score below 40. The third indicator was the Higher Thinking Order Solution (HOTS) which was represented in yellow which had a score of between 50 up to more than 90. The last indicator was Best Practice which was represented by a red-colored line which only had 2 scores namely 0 and 22.

To test heteroscedasticity or to test whether in a regression model there are similar variants or not, a scatter plot graph was used. The scatter plot graph above had 20 training indicators. Each training indicator had 5 different plot forms to represent the HOTS, Andragogy, Training Item, Score, and Prediction.

Behaviour – Conclusion of 5 assessment indicators

The order of the average training indicators was HOTS, best practice, and andragogy. Based on the parametric test, the data was normally distributed and linear.

Behaviour - Best Practice

The graph above shows the constant score of Best Practice on 20 training indicators. 7 of the 20 training indicators had a minimum score of 0, namely impact and implication on situations, decisions, confident, focus communications on the issues, data and information, divides problems for the further analysis, and techniques to solve problems. Based on the parametric test, 7 training indicators were not normally distributed. The remaining 13 training indicators had a maximum score of 22.

Behaviour - HOTS

Higher Thinking Order Solution (HOTS)had a percentage between 50% to 90%. The lowest score was in*communicate effectively with management*of 50 %. The highest score were in *Decisions*, *Focuses communications on the issues,* and *Techniques to solve problems*of 94%. Based on normality and parametric tests, the data was normally distributed.

Behaviour - Andragogy

In the assessment of andragogy on 20 training indicators, the lowest percentage was 3% in*reaction* and *goals.* The highest percentage can be seen in*Analyzes data and information*and *Divides problem for further analysis*of 38%. Based on normality and parametric tests, the data was normally distributed.

Characteristics of Respondents

The graph above shows the characteristics of 306 respondents who provided information about respondents in general related to gender, educational level, position before education and training and involvement in increasing the capabilities of the internal control apparatus before education and training. Based on the graph above, the majority of respondents were male and bachelor while the minority of respondents was doctoral graduates. Based on the position, the majority of respondents were structural officials, then functional officials, and for others that did not exceed 50 people. Based on the graph, 223 respondents were internal control apparatus before education and training.

Description of Variables

1. Improvement of abilities/behavior change (PKP) variable

The graph above shows the improvement of abilities/behavior change (PKP) variable. The improvement of abilities/behavior change after education and training can improve the capabilities of the internal control apparatus which is divided into 17 indicators. The PKP has 5 categories, namely very poor, poor, fair, good, and very good.

Based on the graph above, the majority of respondents stated that PKP was in the good category which was represented by the purple line. Then, the majority of respondents stated that PKP was in the fair category which was represented by the orange line. The 8th and 9th indicators of PKP had a better score than other indicators. In addition, the 8th indicator of PKP with a very good category had lower responses than poor and very poor categories. In fact, the 8th indicator of PKP that was in poor category had a percentage of over 20%.

2. Motivation and Work Environment (MLK) variable

Based on the graph above, majority of respondents agreed with MLK. The 'agree' response was represented by black color with a percentage of over than 50%. Then, the majority of respondents that strongly agreed was represented by purple color. The 'strongly agree' response was in the third place after the 'fair' response in the 5th and 6th indicators of MLK which was represented by orange color. Based on the graph above, 'strongly disagree' response which was represented by the light blue color had the lowest score. So, it can be concluded that most of the MLK indicators were approved.

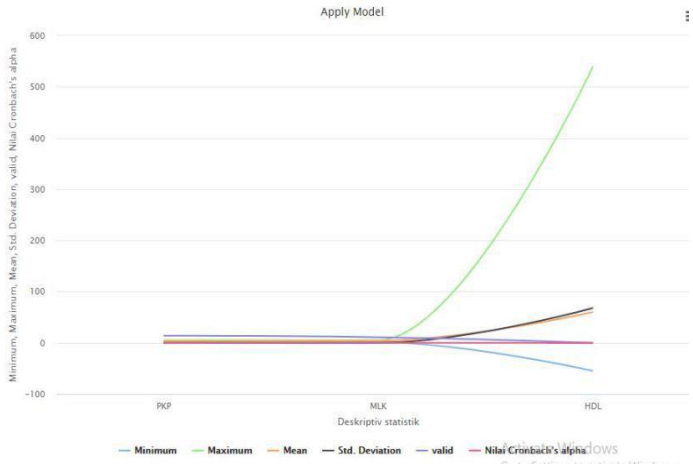
3. Open-Ended Questions

The results of 2 open-ended questions on research variables with 3 answers namely yes, no, and not answer. The first question was on the process of education and training to find out whether it was able to equip respondents with good practices. Then the second question was on the pre/post test to find out the influence on improving the competency of the respondents. Both questions were represented by numbers 1 and 2.

Based on the above graph, 77.1% of respondents answered ‘yes’ on the first question, and 87.3% of respondents answered ‘no’ on the second question. Both questions had 'no' answer of below than 20%.

Descriptive Statistics (Level 3)

Descriptive statistics have a function to describe or give a description of the object under study through sample data or population as they are, without analyzing and making generally accepted conclusions. On descriptive statistics (Level 3), the total data was 306.



1. Descriptive Statistics

The above graph shows minimum (blue), maximum (green), average (orange), and standard deviation (black) scores of research variables. In HDL, the minimum and maximum scores were very different when compared to PKP and MLK. The minimum score of HDL reached -55.56, while the minimum scores of PKP and MLK were 1. The maximum score of HDL reached 540.77 while the maximum scores of PKP and MLK were 5. The standard deviation showed the value of dispersion or average distribution level in a variable and it is useful to know the deviation from the average data series. The greater the standard deviation, the more scatter the data from the average. Conversely, the smaller the standard deviation, the more homogeneous the data. This can be shown by the orange and black graphics which had almost similar movements.

2. Validity Test

The validity test with*Pearson ProductMoment* of 25 indicators showed valid category with purple lineindescriptive statistics level 3 graph.

3. Reliability Test

The reliability test showed that improvement of abilities/behavior change (PKP) variableand Motivation and Work Environment (MLK) variablewith a *Cronbach's Alpha*value of 0.6so all indicators were reliable. This can be seen in pink line in descriptive statistics level 3 graph.

4. Normality Test (4.6 page 68)

Kendall's Tau-b Correlation Test

Based on the above graph, PKP variable had a correlation with MLK because significance level (= 0.000) was lower than 0.05. PKP had no correlation with HDL because the significance level (= 0.145) was higher than 0.05. MLK had no correlation with HDL because the significance level (= 0.859) was higher than 0.05.

5. Interviews and Observation of Work Results

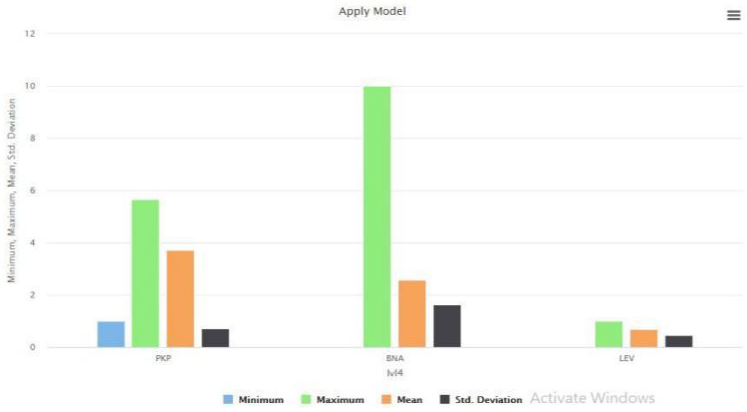
Interviews and observations of work results were carried out on the internal control apparatus which had different numbers of respondents namely inspector of frequency with 1 respondent, inspector of city with 1 city, inspector of Meteorological, Climatological, and Geophysical Agency (BMKG), inspector of Agency for Agency the Assessment and Application of Technology (BPPT), inspector of the Ministry of Agrarian Affairs and Spatial Planning/National Land Agency (ATR/BPM) that each had 4 respondents.

Descriptive Statistics (Level 4) on 171 people

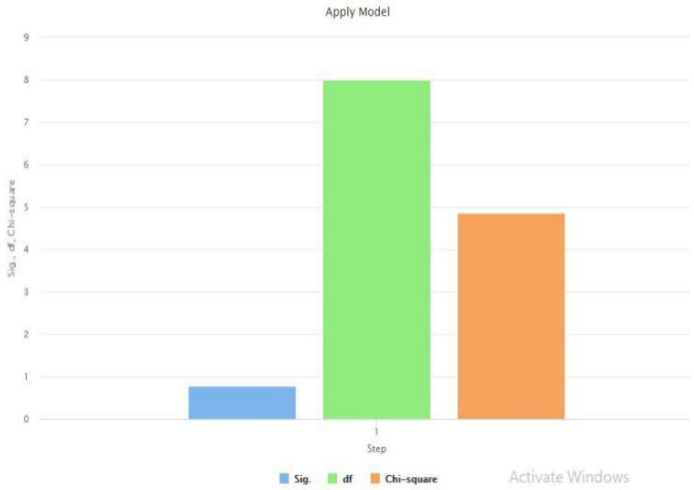
One of the variables at level 4 was obtained from the answer of respondents at Level 3 namely the improvement of abilities/behavior change (PKP) variable that was classified based on the institution of

origin of the respondent (internal control apparatus). The classification results obtained 171 internal control apparatus from the target sample of 226 internal control apparatus or reached 75.63%. The level 4 data tab is presented in appendix 8.

1. Descriptive Statistics

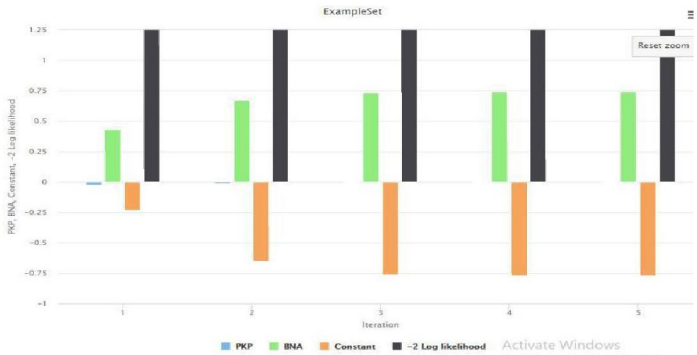


2. Regression Feasibility Test



The graph above shows the results ofregressionfeasbility test with *Hosmer and Lemeshow's Goodness of Fit Test*which was measured by using*Chi-square*. The graph above shows that the value of *Hosmer and Lemeshow's Goodness of Fit Test*(sig.) was 0.771 which was higher than 0.05, so H0 was accepted so that the model was able to predict the value of observations or the model was acceptable because it matched the observational data.

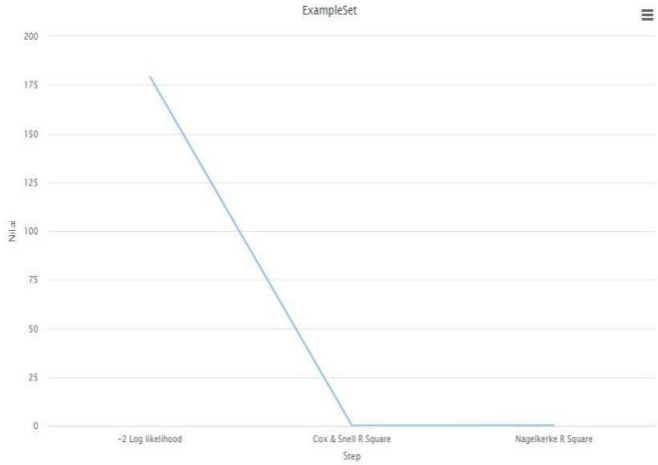
3. Overall Model Fit Test



The graph above shows the results of the overall model fit test which was used to assess whether the model is in accordance with the data. Based on the graph above, -2 log likelihood was equal, but not the same. Then the coefficient constant had a minus value which was getting bigger in each step from 1 to 5. In PKP graph, steps 1 and 2 had minus values. Steps 3, 4, and 5 had unclear values of PKP because the values was

close to 0 namely -0.007. The BNA value was represented by green with a value that was over than 0 on all five steps.

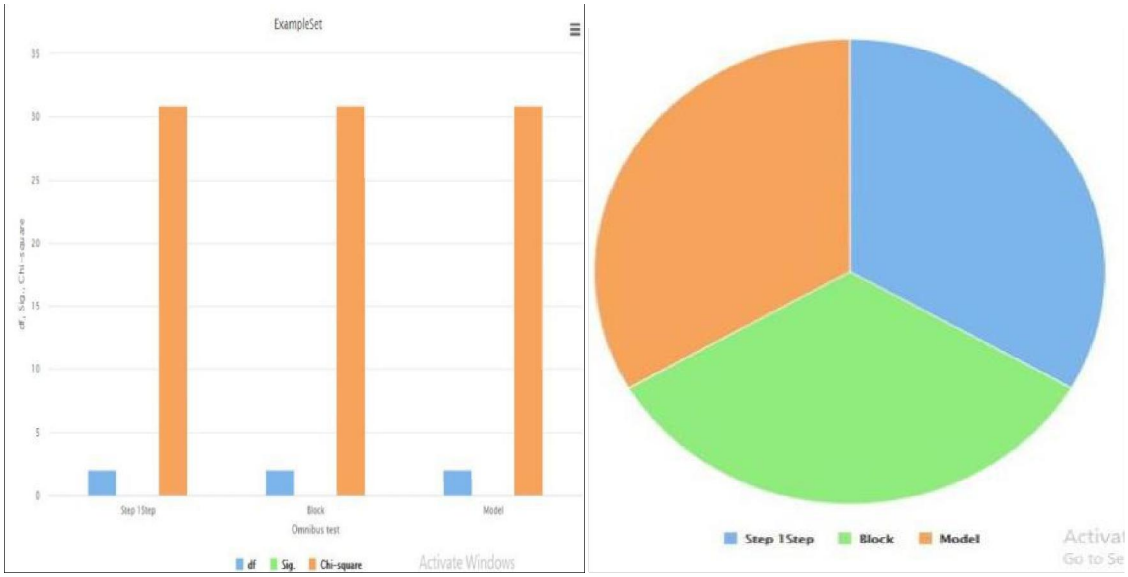
4. Coefficient ofDetermination Test



The graph above shows the coefficient of determination test with Nagelkerke's R square. The purpose of this test was to determine the combination of independent variables, namely the improvement of abilities/behavior change and coaching to explain the dependent variation, namely improving the level of capability of internal control apparatus.

Based on the graph above, -2 log likelihood reached 179.212[a]. Cox &Snell R square and Nagelkerke R square had a value of over than 0 but below 1.

5. Simultaneous Test

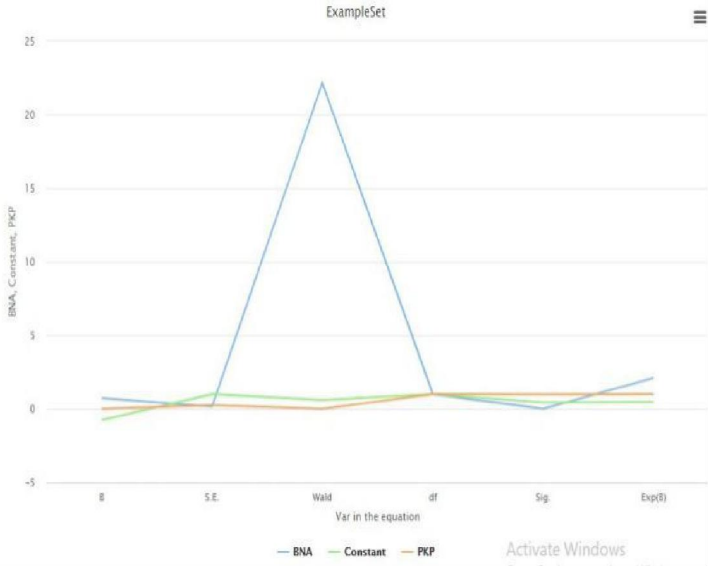


This test 2 qs conducted to test whether the independent variables consisting of improvement of abilities/behavior change and coaching simultaneously had an influence on the improvement of capability of internal control apparatus. Hypothesis testing was performed by comparing the probability value (sig) represented in green, with a significance level (α).

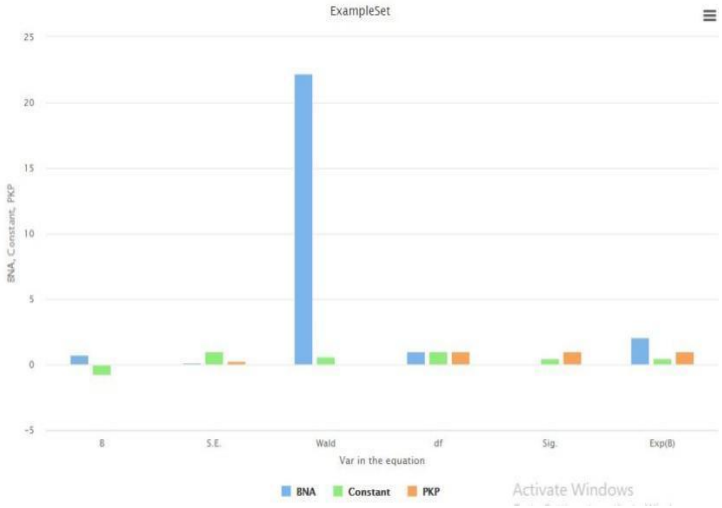
The results of the graph above, simultaneous testing with the omnibus test, namely step 1, block and model showed a sig = 0,000 with a value of less than 0.05 so H0 was rejected which means the improvement of abilities/behavior change and coaching had an influence on the improvement of capability of internal control apparatus.

The circle graph shows that the chi-square, df, and sig of step 1, block, and the model had the same constant value.

6. Partial Test



The graph above shows the partial test on variables which consists of B, S.E., Wald, df, Sig., Exp (B). The highest score of BNA was represented by blue at the point where the Wald variable was reached 22,177. Meanwhile, values below 0 occur at a constant line represented by green at point B which was -0.768.



The partial test showed thatimprovement of abilities/behavior change had sig=0.979 which was higher than 0.05 so that H0 was accepted.So it can be concluded that at 95% significance level, PKP had no influence on the improvement of capability of internal control apparatus The coaching variable had sig=0.00 which was lower than 0.05 so that H0 was rejected.So it can be concluded that at 95% significance level, coaching had an influence on =the improvement of capability of internal control apparatus.

IV. CONCLUSION

Based on the research results above, it can be concluded that the instructor and implementation needs to be improved to achieve the desired work target. In addition, the education and training must be improved, so that good quality of supervision can be achieved.

REFERENCES

[1] Notoatmojo, Soekidjo, Pengembangan SDM (Jakarta: RinekaCipta, 2017), p. 60
[2] Alqur'an dan Terjemah, Kementrian Agama RI, DirektoranJendralBimbingan Masyarakat Islam Indonesia:PT Adhi AksaraAbadi , 2015
[3] Kotirde&YunosdalamIntan, supervision, p. 38.

[4] Mette, Anderson, Nieuwenhuizen, Range, Hvidston, dan Doty in M. Nuh, Cross Sectional (Jakarta: PT Rajawali, 2017), p. 87
[5] Bagiyono, EvaluasiKualitas dan EfektifitasPelatihan, Proceeding Seminar Internasional, SDM, 2014, p. 97-103
[6] Arikunto et al, Evaluasi Program Pendidikan,PanduanTeoritisPraktisBagiPraktisi Pendidikan. BumiAksara, Jakarta, 2017, p. 65
[7] Pusdiklat reinforcement Kompetensi supervisor Sekolah, Evaluasi Efektifitas Pelatihan, dilanjutkan dengan Proceeding Internasional, 2016, p. 92