AN EVALUATION MODEL OF EDUCATIONAL QUALITY ASSURANCE AT JUNIOR HIGH SCHOOLS

*[1]Sugiyanta; [2]Soenarto
[1]Educational Quality Assurance Institution, Kalasan, Tirtomartani, Sleman, Yogyakarta 55571, Yogyakarta, Indonesia
[2]Faculty of Engineering, Universitas Negeri Yogyakarta, Jl. Colombo No. 1, Depok, Sleman, 55281, Yogyakarta, Indonesia

Abstract
The study was to develop an appropriate evaluation model of quality assurance (QA) for evaluating the programs of the educational QA (EQA) at junior high schools. The study was a research and development study that referred to the steps developed by Borg and Gall. The results of the study show that the evaluation model of EQA in junior high schools consist of the implementation of QA system and the performance of QA. The constructs for the instrument of QA system implementation consisted of planning, implementation, monitoring and evaluation, and the act of revision is based on the exploratory factor analysis at the significance level of 0.000. The constructs for the instrument of EQA performance consisted of: resource development; program and activity development; participation, satisfaction, knowledge change, attitude change, and behavior change of school community; social, economic, and school environmental development based on the exploratory factor analysis at the significance level of 0.000. The feasibility of the evaluation model is in a good category based on experts', users', and practitioners' judgment and the evidence found in the field testing.

Keywords: *evaluation model, quality assurance, junior high school*

*Corresponding Author.
e-mail: pakdhetato@yahoo.co.id

Introduction

The Government Regulation Number 19 Year 2005 regarding the National Education Standards mandates that each educational provider or each school is compulsory to implement the quality assurance (QA). The educational quality assurance (EQA) is a sequential process of determining and meeting the standards that have been designed and implemented consistently and sustainably in order to gain trust. The implementation of QA is intended in such a way that the schools meet or exceed the National Education Standards (*Standar Nasional Pendidikan*, SNP) that have been stipulated.

An evaluation of the implementation of QA is necessary for the decision taking, the policy making and the subsequent program designing. The measurement and the evaluation in the QA of school are conducted in order to attain information regarding the efforts to meet the quality reference and to answer the questions of school performance in displaying the quality commitment through the clear, well-planned and measured mechanism (Loder, 1990, p.5).

Based on the preliminary study in 2014 of the senior high schools in the Province of Yogyakarta Special Region, the researchers attain information that schools have not implemented the self-evaluation in relation to the implementation of the evaluation of the QA that has been conducted. In addition, there has not been any type of evaluation that is able to provide the information regarding the performance of the quality assurance.

Based on the above background, the measurement and the evaluation of the efforts to QA should be implemented through a system of QA and the measurement of school performance in achieving the quality assurance. The efforts to meet the quality standards could be conducted through the implementation of QA system; according to Edmond (1979, pp.15-27), the QA system consists of the consistent and effective planning and implementation strategy. In addition to this aspect, there are two other components namely monitoring and evaluation and improvement measure. The aspect

of QA performance, as proposed by Loder (1990, pp.189-200), could be developed based on the internal and external needs, for example: curriculum input, teachers, clarity of school programs in meeting the quality standards, available resources, participation of school community in improving the quality standards and the effect of these aspects on the surrounding environment. These aspects are very important because quality is not static but, instead, is dynamic in accordance with the development of the environmental and the societal needs.

Evaluation Model of Educational QA
(*Evaluasi Penjaminan Mutu Pendidikan*, EPMP)

The EPMP Model is a combination of the discrepancy model by Provus and the hierarchy model by Rockwell and Bennett (2004, pp.5-7). With several omissions and expansions in several aspects of evaluation. In details, the evaluation is directed toward two aspects namely the level of evaluation and the scope of evaluation. On the other hand, the scope of evaluation covers the implementation of QA system and the QA performance. These statements have been in accordance with the study by Widoyoko (2013, pp.41-54), which states that educational evaluation by means of process approach will be more comprehensive so that the educational evaluation will generate more complete information. The focus of evaluation of the performance of QA is based on seven items of QA performance namely: (1) Resources development; (2) programs and activities development; (3) target audience participation; (4) client satisfaction or reaction; (5) knowledge, attitude and skills change; (6) school communities' behavior change; and (7) social, economic, and school environmental development of the schools.

As shown in Figure 1, the evaluation of the implementation of QA system could be seen from the four main stages in the quality assurance, namely: (1) The QA program planning that consists of the human resources preparedness, the program plan; the clarity of quality standards and the implementation procedures; (2) the program implementation that consists of QA activities realization, the
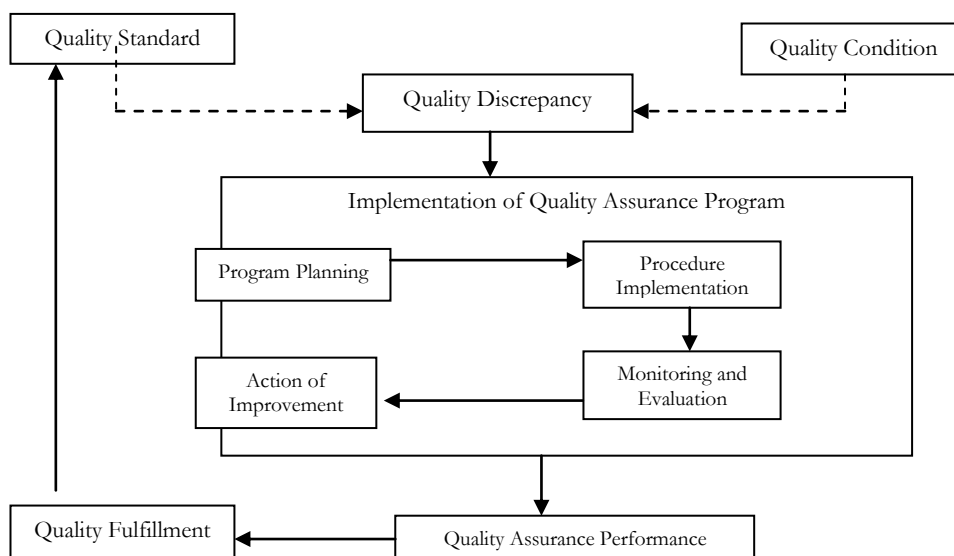
Figure 1. The Evaluation Model of Educational Quality Assurance

compliance toward the quality procedures and the relevance between the actions and the procedures; (3) the monitoring and evaluation that consists of the monitoring and evaluation program possession and the evaluation, implementation and reporting program; and (4) the improvement action that consists of the action plan design, the action plan implementation and the action plan results.

Furthermore, to attain the information regarding the two aspects that have been mentioned, self-evaluation type study was implemented. The reasons for developing the evaluation model of QA in the study are as follows. First, the focus of evaluation is the process of meeting the quality standards; therefore, the results of evaluation are expected to provide a description of the school's profile and efforts to meet the quality promises. Second, the focus of evaluation is to measure the school's performance in implementing the educational QA system. Third, the evaluation model of QA that that has been developed is in accordance with the cycle of educational quality assurance.

The scope of the study is the following research questions. (1) What should be the appropriate evaluation model of internal QA for evaluating the implementation and the performance of QA at junior high schools? (2) What should the model of indicator constructs be in order to be implemented for measuring the implementation of internal QA

system at junior high schools? (3) What are the indicator constructs that could be implemented for measuring the performance of internal QA at junior high schools? (4) What is the model feasibility of internal QA at junior high schools like? Those problems are proposed to gain more comprehensive evaluation results on how the schools implement the quality assurance system and how good the schools' performance is in achieving the quality through such QA program.

In relation to the scope of the study, the objectives of developing the product in the form of an evaluation model are as follows: (1) To generate an appropriate evaluation model of internal QA for evaluating the program of educational QA at junior high schools; (2) to generate the model of indicator constructs that could be implemented for measuring the implementation of an educational QA system at junior high schools, which reflect the efforts of quality fulfillment conducted by the schools through well-planned, systematic, and sustainable steps of quality assurance; (3) to develop a model of indicator constructs that could be implemented for measuring the performance of educational QA at junior high schools, which may provide information on the results of the implementation of the schools' QA program; and (4) to attain information regarding the level of educational QA at junior high schools.

The advantages of the product development are as follows: (1) Developing an evaluation model that could be implemented for evaluating the program of educational QA within the schools; (2) providing an alternative model that might generate information regarding the QA of schools; and (3) being able to be implemented by both the schools and the general public for performing self-evaluation in the level of educational quality assurance.

The Criteria of the Level of Educational Quality Assurance

The criteria developed in the study employ mutual-adaptive approach which is the combination of pre-ordinary approach and fidelity approach. The evaluation criteria are designed before the field implementation. These criteria are developed based on theories of QA and the perspective of the program developers as well as the characteristics of the program in the field. Therefore, the criteria development in the model development is the Criterion-Referenced Evaluation, in which the results of the evaluation are based on the tasks themselves and not on the performance of typical people.

In relation to the needs, within the model development, the criteria of QA are based on the objectives and tasks of schools within the QA in accordance with the quality standards that have been stipulated by the respective schools as having been explained in Figure 1.

The level of educational QA in the evaluation model of QA performance within schools involved in the model development has four performance levels that have been developed based on the theory and the characteristics of educational QA program in the schools. The criteria of QA performance are as follows:

Level 4 = The educational quality has been totally assured
Level 3 = The educational quality has been assured
Level 2 = The educational quality has been less assured
Level 1 = The educational quality has been totally less assured

Those levels of QA performance are gained based on the average score of the assessment results of the whole aspects of QA system implementation and performance.

Level 4 is the highest level of the QA and the meaning of attaining the Level 4 is that the school or the school has very clear and measurable quality standards, performs the efforts to meet the quality standards seriously and totally comply to the quality standards so that the quality has been very controlled and very developed sustainably. Thereby, the educational quality is very assured/could be assured.

Level 3 refers to the level in which the school or the school has not attained the clear quality standards, affords to meet the quality standards seriously under the clear procedures, and has conducted the procedures most of the time. However, the school or the school has not performed the tight control. In general, the school or the school might implement/assure most of the quality standards that have been stipulated.

Level 2 refers to a condition in which the school or the school has not defined the clear quality standards, has procedures that have been formally stipulated and that have been part of mere habit, and does not pay sufficient attention to the quality control. As a result, the quality becomes less assured.

Level 1 refers to a condition in which the school or the school does not have clear quality standards, does not have procedures for meeting the quality standards, and does not have internal quality control. As a result, the quality becomes totally not assured.

Method

The study was a research and development that adopted the stages proposed by Borg and Gall. The study was conducted in Yogyakarta from May 2012 to October 2014. The number of subjects in the first experiment, the second experiment, and the operational experiment was increasing. For the data gathering method, checklist method and the questionnaire distribution were implemented. The instrument development was conducted through the following steps: (1) Determining the constructs, namely creating the limitations

regarding the variables to be measured based on the in-depth theoretical review; (2) defining the factors that defined the elements in the constructs; and (3) designing the items that described each factor in the form of statements.

The instrument validity was based on the content and the constructs. For the test of content validity, expert judgment through forum group discussions (FGD) was conducted by looking for considerations from the people who understood more about the substance of the study namely the quality assurance. The construct validity of the instrument was analyzed by means of Exploratory Factor Analysis (EFA). Then, the reliability of the instrument was calculated by means of the Alpha Cronbach technique by adopting the requirements proposed by Kaplan and Saccuzzo (1982, p.106); if the alpha coefficient is bigger than 0.7, then the instrument is considered reliable.

Data Analysis Technique

The quantitative analysis was conducted to analyze the fitness of the measurement model. Then, the qualitative analysis was directed toward analyzing the evaluation model and the evaluation guideline that was developed. The quantitative data analysis, the results of respondents' assessment toward the model, the instrument and the evaluation guideline were converted into the qualitative criteria that were implemented. Next, the regulations developed by Sudijono (2003) as presented in Table 1 were referred in order to determine the feasibility of the development model.

Table 1. The conversion of the quantitative data into the qualitative data

Interval	Classification
$X > Xi + 1.8.Sbi$	Very Good
$Xi + 0.6.Sbi < X \leq Xi + 1.8.Sbi$	Good
$Xi – 0.6.Sbi < X \leq Xi + 0.6.Sbi$	Moderate
$Xi – 1.8.Sbi < X \leq Xi – 0.6.Sbi$	Poor
$X \leq Xi – 1.8.Sbi$	Very Poor

Notes :

Xi (Ideal Mean) = ½ (Ideal maximum score + Ideal minimum score)

Sbi (Ideal standard deviation) = $^1/_6$ (Ideal maximum score – Ideal minimum score)

X = Empirical score

The model testing was conducted by implementing several indicators. The first indicator was (Kaiser-Meyer-Olkin) KMO; if the KMO was bigger than 0.5 then the data could be analyzed further (Ghozali, 2009, p.394). The second indicator was the size of load factor that displayed the size of item load of its factors. The criteria implemented, as proposed by Tabachnick and Fidel (1983), showed that the factor load was bigger than 0.71 (very good); 0.63 (very good); 0.55 (good); 0.45 (moderate); 0.32 (poor). For the item load, the benchmark was 0.55; if the item score was bigger than 0.55 then the item could be implemented. The third indicator was the Eigen value whose score should be bigger than 1 (one); this score was the overall representation in relation to the relevance among the factors that were selected as the indicators. If the cumulative percentage was bigger than 50%, the research might state that the factor selection was appropriate.

The data gathering was conducted by means of a document checklist and questionnaire. The subjects in the first experiment were 41 respondents from the QA team members, teachers, employees, school committee members, and the students. These respondents came from three schools. The second experiment involved 120 respondents from five schools. Last but not least, the operational/implementation experiment involved 258 respondents from 10 schools.

Findings and Discussion

Results of Product Testing

Based on the theoretical studies and the preliminary studies conducted at three junior high schools, there are 12 essential indicators in the implementation of internal QA system at the schools. This result was then consulted with experts, educational practitioners in a forum group discussion (FGD). At the preliminary stage, the FGD was conducted in order to attain the feedbacks both in terms of evaluation procedures and evaluation model preliminary design that took the constructs of QA evaluation, instrument form, data source and data gathering method and evaluation procedures. From the results of assessment

toward the indicators, it was found 12 indicators with very important category and four indicators with important category for the implementation of educational QA items. On the other hand, for the aspect of QA performance, it was found 13 indicators in a very important category and six indicators in an important category. The draft of the model was then experimented gradually.

First Experiment

The first experiment was conducted in order to gain feedbacks from the practitioners and the users of QA evaluation instrument at schools regarding the evaluation model feasibility. The educational QA evaluation instrument was the instrument resulted from the revision and the preliminary draft resulted from the preliminary review and the FGD that was validated by the experts. The first experiment was conducted at three schools, namely State Junior High School of 1 Banguntapan, State Junior High School of 3 Tempel and State Junior High School of 1 Berbah. The respondents in the first experiment were the Team of School Development (*Tim Pengembang Sekolah*, TPS) or the Team of QA consisting of the principal, teachers, employees, school committee members, students, and students' parents.

Quantitatively, from the results of model and feasibility instrument and the guidelines of questionnaire evaluation from the first experiment, it was found that: (1) the mean score of the evaluation model assessment was equal to 3.6 (low category); (2) the mean score of the instrument clarity assessment was equal to 3.653 (low category); and (3) the mean score of self-evaluation assessment was equal to 3.615 (low category). Based on these results, the elements in each assessment were improved and the second experiment was conducted.

Second Experiment

The second experiment was the main field experiment whose objective was to gain feedbacks from the wider field especially in relation to the evaluation model and evaluation instrument. The analysis of the fitness of measurement model was conducted toward two evaluation instruments namely: (1) the QA system implementation instrument and (2) the QA performance instrument. Then, the analysis of the fitness of the measurement model was conducted by implementing the exploratory factor analysis (EFA) technique with the assistance of SPSS 17.0 program.

Based on the results of the analysis of the QA system implementation instrument, it was found that the KMO score was equal to 0.551 at the significance level of 0.000. The score implied that the data could be analyzed further. The KMO score was bigger than the required score (0.50). Therefore, once again, the data were analyzed further. Then, based on the multivariate correlation test with Bartlett, it was found that Sig 0.000 was smaller than the Alpha 0.05; as a result, the researcher could conclude that there was correlation among the multivariate. In other words, the data in the second experiment was feasible for further analysis. The size of correlation value among the multivariate variables, based on the coefficient of Measure of Sampling Adequacy (MSA) in the Anti Image Correlation, showed that almost all items in all variables were bigger than 0.5; as a result, the variables could be predicted and could be analyzed further (Santoso, 2014, p.69).

The number of variants and variables that could be explained by the factors that were designed could be seen from the communality value. From the results of the analysis, it was found that 40 items of the QA system implementation assurance and the 40 items had communalities score that was bigger than 0.50. In other words, the variables within the evaluation model could be explained by the factors that were explained. The total cumulative variance from the results of the analysis was 77.514%, which implied that the variables in the study might explain 77.514% of nine factors that were designed with various item distributions.

From the results of the testing of the QA performance instrument, it was found that the KMO score was equal to 0.860 at the significance level of 0.000 and, therefore, the data could be analyzed further. The results of the testing by means of multivariate correlation with Bartlett showed that the alpha

score was equal to 0.000, which was smaller than 0.05. Therefore, the researcher could conclude that there was a correlation among the multivariate variables. The results of inter-multivariate variable correlation showed that the coefficient value of Measure of Sampling Adequacy (MSA) in the Anti Image Correlation in almost all items within all variables was bigger than 0.5; therefore, the researcher could conclude that these variables could be predicted and could be analyzed further.

The number of variants from the variables that could be explained by nine factors that were designed could be seen from the communality value. From the 55 items, there were 50 items or 90.90% items of the QA performance instrument whose communality value was bigger than 0.50. The finding implied that the variables within the evaluation model could be explained by the factors that were designed. The total cumulative variance from the results of the analysis was equal to 87.995%, which implied that the variables in the study might explain 87.995% of the factors that were designed.

Operational Experiment

The operational experiment was conducted after the researcher improved the items that were not relevant to the content and the factors that were designed; as a result, the items were grouped into four factors in accordance with the theoretical model. In this stage, the experiment involved the subjects in a wider scale, namely 10 schools that were located in four counties and one municipality of the Province of Yogyakarta Special Region. Twelve respondents were selected from each school, and the respondents consisted of principals, teachers, employees, students, and parents/foster parents. Overall, the third experiment involved 128 respondents consisting of principals, school QA team members, teachers, and employees, 130 respondents consisting of the school committee members and the parents in order to gain additional information.

Based on the results of the experiment, it was found that the KMO score was bigger than the required KMO score (0.50). Such coefficient value belonged to the Meritarius or beneficial category so that the data could be analyzed further. In addition, based on the multivariate correlation test with Bartlett, it was apparent that Sig 0.000 was smaller than the Alpha 0.05. Thereby, it can be concluded that there was a correlation among the inter-multivariate variables. As a result, the data were feasible for further analysis. In addition, the communality value showed that the number of variants from the variables could be explained by the four existing factors. The inter-multivariate variable correlation score, based on the coefficient of Measure of Sampling Adequacy (MSA) in the Anti Image Correlation, in all items from all variables, was bigger than 0.5 and, therefore, it can be concluded that these variables could be predicted and be analyzed further.

The number of variants from the variables that could be explained by the four factors that could be explained was apparent from the communality value. Based on the results of the analysis, there were nine items whose communality value was under 0.50, namely: the planning factor, items 3, 14 and 15; the implementation factor, items 18 and 21; the monitoring factor, items 31, 32 and 37; and the action factor, item 47. Meanwhile, the communality value of the other 39 items was bigger than 0.50.

The cumulative percentage of the analysis results for the four factors was quite good, namely 56.526%. The cumulative percentage showed that the instrument might explain the factors in the QA evaluation model for about 56.526%. The percentage had met the requirements proposed by Tabachnick and Fidel (1983) that if the cumulative percentage was bigger than 50% then factor selection would be appropriate. The four existing factors had the Eigen value > 1, which showed that the selected factors were used as the indicators of a characteristic or a trait. Thereby, it can be concluded that there were four factors existing in the constructs of the educational QA system implementation instrument, namely planning, implementation, monitoring, and evaluation and the action of improvement could be explained by the variables that were observed within the study.

The size of the item load of the factor was shown by the size of the factor load from each variable in the component matrix table. Based on the results of the component matrix rotation, there were still six items whose correlation coefficient was under 0.55.

Based on these results, it can be concluded that 42 out of 48 items (87.50%) of the instrument had a factor load which was bigger than 0.55. The finding implied that the item load of the factor was good so that in general the items within the instrument were valid and could be used. On the other hand, the items that were not valid were omitted or were not put into the final product of the evaluation instrument.

The above results of the analysis showed that the QA system implementation instrument had four factors, namely planning, implementation, monitoring, and evaluation and the action of improvement was proven by the size of the factor from each factor variable that was bigger than 0.50. The results of Rotated Component Matrix also proved that the instrument items were concentrated according to the factors that were hypothesized theoretically.

The reliability of the EQA system implementation had increased in terms of the reliability coefficient value from the first experiment and the operational experiment.

Table 2. The recapitulation of the reliability coefficient for the QA system implementation instrument

No.	Name of Factor	Coefficient of Cronbach Alpha	
		First Stage Test	Final Test
1	Planning	0.937	0.940
2	Implementation	0.899	0.906
3	Monitoring and Evaluation	0.927	0.930
4	Action of Improvement	0.861	0.872

The testing of the constructs of the QA performance instrument consisted of a latent variable with seven observed variables. The statement was supported by the results of factor analysis. The results of the analysis showed that the KMO score of the QA performance variables was equal to 0.785. The

score was bigger than the required score (0.50) and, as a result, the analysis belonged to the Meritarius or beneficial category. Based on the score, the data on the QA performance could be analyzed further. The multivariate correlation test by means of Bartlett and Sig. also showed that p=0.000. The score was smaller than 0.05; therefore, it can be concluded that there was inter-multivariate variable correlation. In other words, the data from the third experiment could be feasible for the further analysis. The size of Measure of Sampling Adequacy (MSA) coefficient in the Anti Image Correlation in all items for all variables was bigger than 0.5. In relation to the statement, it can be concluded that these variables could be predicted and analyzed further (Santoso, 2014, p.69). Last but not least, the communality value also showed that the number of variants from the variables could be explained by seven existing factors.

From the results of the analysis, it was found that from 55 items, there were five items whose score was lower than 0.50. The five items were found in the following factors: resource development factor, item 6; program development factor, item 22; school community satisfaction factor, items 36 and 37; school community behavioral change factor, item 52. On the other hand, the score of the other items was higher than 0.50. These findings implied that the factors of the educational QA performance were: (1) resource; (2) program and activity development; (3) people participation; (4) customer satisfaction; (5) knowledge, attitude and skill change; (6) behavioral change; and (7) social, economic and environmental development. All of these factors could be explained by the variables that were observed in the study.

Based on the above results, it could be stated that factors 3, 4, 6 and 7 had valid observed variables. Then, the invalid items were omitted and they were not included in the instrument.

Based on the above results, it could be stated that 42 out of the total instrument items had a load factor that was bigger than 0.55. This implied that the item load of the factor load was good so that in general the items in the instrument were valid and could

be used. Furthermore, for the sake of improving the instrument, the invalid items were omitted and were not put in the final product of the evaluation instrument.

The reliability of the educational QA performance instrument from the first experiment to the final stage experiment was increasing. In details, the alpha coefficient from each factor in the QA performance instrument is presented in Table 3.

The results of the final stage experiment showed that each factor in the evaluation instrument showed that all of the factors had the Alpha Cronbach coefficient that was lower than 0.7. From all of the items in the seven factors under analysis, all of the items had a strong correlation with each factor that was hypothesized. Thereby, it can be stated that all of the variables in the factors was quite reliable and all of the items from each factor had high reliability level so that these items could be used as the items in the educational QA performance instrument.

The Model Feasibility and the Instrument Clarity

The aspect of model practicality in the first experiment gained many criticisms from the schools because the instrument was too complicated for the evaluation implement-ation since the evaluation involved many parties, starting from the school committee members, employees, students and parents. In addition, within the data analysis, the results of the evaluation were considered compli-cated if the evaluation was implemented manually. Table 4 presents the assessment results of the evaluation model feasibility.

The results of the improvement showed the increase on the model practicality level as shown by the results of the evaluation in the second experiment and operational experi-ment; in both experiments, the results of the evaluation was in 'good' category. Thereby, the model of the educational QA evaluation could be implemented without any revision.

The Assessment of the Evaluation Instrument Clarity

As seen in Table 5, in the first experi-ment, from five aspects of the instrument clarity, there were two aspects that were good and three aspects that were not good. The aspects that were good were the criteria clarity and the instrument manual clarity. The two aspects did not need any fundamental revision.

Table 3. The recapitulation of the reliability coefficient for the QA performance instrument

No	Name of Factor	Coefficient of Cronbach Alpha	
		First Stage Test	Second Stage Test
1	Resource Development	0.870	0.980
2	Program and Activity Development	0.967	0.973
3	School Community Participation	0.843	0.843
4	School Community Satisfaction	0.891	0.896
5	Knowledge, Attitudes and Skills Change	0.930	0.930
6	School Community Behavioral Change	0.772	0.838
7	School Social, Economic and Environmental Development	0.855	0.855

Table 4. The model feasibility

No	Aspect of Assessment	Score and Category		
		First Experiment	Second Experiment	Operational Experiment
1.	Evaluation model clarity	3.6667 (Good)	3.9375(Good)	3.8788(Good)
2.	Evaluation procedure clarity	3.4667(Moderate)	3.8125(Good)	3.8030(Good)
3.	Evaluation model practicality	3.4667(Moderate)	3.8125(Good)	3.8030(Good)
4.	Evaluation model benefits	3.8000 (Good)	4.2083(Good)	4.0606(Good)
5.	Model use feasibility (time, cost and efforts)	3.6667 (Good)	3.8125(Moderate)	3.8182(Good)
	Average	3.6133(Good)	3.9427(Good)	3.8727(Good)

The aspect of indicator clarity, of statement item readability, and of font shape and size relevance was not good. The improvement that was done in the first experiment was the one on the aspect of indicator clarity and of statement item readability. The two aspects were designed in a simpler manner and were adjusted to the respondents' characteristics. In addition, the fonts were designed in an ordinary layout so that the fonts would be more readable.

The results of the improvement showed good results in the second experiment and the operational experiment, and the results were in 'good' category. Based on the results of the final experiment, the evaluation instrument could be stated as feasible as the instrument of educational quality instrument evaluation at schools.

The Assessment of the Evaluation Manual

The assessment of the evaluation manual feasibility was conducted from the first experiment. Based on the results of the assessment as shown in Table 6, the results that were not in the good category were improved. The results of improvement in the first and second experiments show improvement on the quality of the evaluation manual.

The Final Product Review

The model of the educational QA evaluation was an evaluation model for the process of national education standards fulfillment through the implementation of educational QA system and the educational QA performance.

Based on the results of the experiment and the analysis of the evaluation and implementation at schools, the evaluation model of the educational QA was fit into the evaluation of the implementation of QA and performance system within the schools in terms of implementing the educational QA programs.

The evaluation model of the educational QA was an evaluation model for the process of national education standards fulfillment through the implementation of the educational QA system and the educational QA performance.

Based on the results of the experiment, the analysis and the implementation of the model in the schools, the evaluation model of the educational QA was fit into the evaluation on the implementation of the QA and performance system within the schools in terms of implementing the educational QA programs.

Table 5. The assessment of the instrument clarity

No.	Aspect of Assessment	Score and Category		
		First Experiment	Second Experiment	Operational Experiment
1.	Criteria clarity	3.6000(Good)	4.0000(Good)	3.9697(Good)
2.	Instrument manual clarity	3.6000(Good)	4.0625(Good)	4.0455(Good)
3.	Indicator clarity	3.7333(Worse)	4.0625(Moderate)	4.2121(Very Good)
4.	Statement item readability	3.6000(Worse)	4.1458(Good)	4.1515(Good)
5.	Relevance between the font size and the font shape	3.7333(Worse)	4.1667(Good)	4.2121(Very Good)
	Average Score	3.6533(Worse)	4.0875(Good)	4.1182(Good)

Table 6. The Assessment of the Evaluation Manual

No.	Aspect of Assessment	First Experiment	Second Experiment
1.	Relevance between the content and the development model	3.60 (Good)	4.0208 (Good)
2.	Manual clarity	3.60 (Good)	4.0417 (Good)
3.	Evaluation stage clarity	3.73 (Good)	4.0833 (Good)
4.	Recommendation direction and objective of evaluation results clarity	3.60 (Good)	4.1042 (Good)
5.	Sentence comprehensiveness	3.73 (Good)	4.1250 (Good)
	Average Score	3.65 (Good)	4.0750 (Good)

The followings are the characteristics, the strengths and the weaknesses of the evaluation model of educational QA that is developed within the study.

The Model Characteristics

According to the evaluation objectives, there are several characteristics of the evaluation model of the educational QA that distinguish this evaluation model from other evaluation models, namely: (1) the evaluation model is implemented in order to evaluate the process of fulfilling the national education standards or the standards that are stipulated by the schools through the implementation of the QA system and of QA performance; (2) the evaluation model could be implemented by the schools, the school supervisors, the public and the related parties to identify the level of the schools' QA in fulfilling the standards implemented; (3) this evaluation model consists of two aspects, namely the QA system implementation and the QA performance; (4) this evaluation model has the criteria with four levels of QA that represent the level of the school's QA in the process of fulfilling the standards stipulated by the schools; and (5) this evaluation model is open and transparent or, in other words, the data gathering is conducted openly by involving all components of school community and the results and the recommendations of improvement are presented transparently. These criteria show the effectiveness of the schools' QA program and the schools' commitment in fulfilling the quality promises.

The Model Strengths

The evaluation model has several strengths, namely: (1) The evaluation results could be directly implemented for improving the schools' management in implementing the QA programs because the recommendation of improvement is based on the data of QA implementation; (2) the evaluation model could be implemented for portraying the process of fulfilling the national education standards or the standards stipulated by the schools; and (3) the evaluation model is equipped with the recommendation formats, the school budget plan/the school budget plan and the formats of evaluation results report, which are very useful for the schools in performing the follow up of the evaluation results.

The Model Weaknesses

The developed evaluation model in the study still has several weaknesses, namely: (1) the evaluation model development is limited to the Province of Yogyakarta Special Region; (2) the model implementation is limited to the schools that have applied the educational QA programs; (3) the data sources are limited to the documents and the respondents from the school community and do not involve the members of surrounding communities; (4) the evaluation model is not able to portray the overall quality culture, which is the final objective of quality assurance; (5) the aspects under evaluation are still limited to the main components of quality assurance; and (6) the questionnaire and the document checks of the QA are still limited to the information provided by the respondents and the documents and are not equipped with an observation instrument.

Conclusions and Recommendations

Conclusions

Based on the data description and the data analysis, the following conclusions can be drawn: (1) the appropriate evaluation model for evaluating the educational QA of junior high schools consists of an evaluation of the educational QA system and an evaluation of the educational QA performance; (2) the constructs of educational QA system implementation instrument consist of four dimensions, namely planning, implementation, monitoring, and evaluation and the action of the improvement based on the exploratory factor analysis and the factor load of all variables is bigger than 0.50 and belongs to the good category; (3) the constructs of the educational QA performance instrument consist of seven dimensions, namely resource development, program and activity development, school community participation, school community satisfaction, knowledge, attitude and skills change, school community behavioral change

and social, economic and environmental change, and based on the exploratory factor analysis the loading factor of all variables is bigger than 0.50 and belongs to the good category; and (4) the feasibility of the evaluation model for the educational QA at junior high schools belongs to the good category based on the expert validation, the user validation, and the practitioner validation as well as the evidence found in the field study.

Recommendations

Based on the conclusions, the following recommendations can be proposed: (1) the evaluation model (EPMP) could be implemented as an alternative both for the schools in implementing the self-evaluation and for the related parties such as the Office of Education, the Institution of Educational QA (*Lembaga Penjaminan Mutu Pendidikan*, LPMP) and Centre for Development and Empowerment of Teachers and Education Personnel (*Pusat Pengembangan dan Pemberdayaan Pendidik dan Tenaga Kependidikan*, P4TK) in viewing the efforts of fulfilling the quality/standards by measuring the implementation of the QA system and performance; (2) the results of evaluation could be turned into the matters of managerial supervision for the school supervisors especially at junior high schools; (3) the QA evaluation could be developed online so that the time and the space limitation could be minimized; (4) the development of the QA system implementation is limited to the four main components of QA system and, on the other hand, the QA performance is still limited to the seven dimensions according to the exploratory factor analysis, and as a result, these findings provide an opportunity to develop an evaluation model of QA that will be more comprehensive; (5) the QA system implementation instrument and the QA performance instrument are still limited to the questionnaire and the document check on the implementation, and therefore, the future researchers might develop the more comprehensive instruments; and (6) the data source in the evaluation model is still limited to the school community and the document, so

future researchers might develop an evaluation model of QA with more representative and variable data sources.

References

Borg, W.R. & Gall, M.D. (2003). *Educational research: An introduction*. London: Longman.

Edmond, R.R. (1979). Effective school for the urban poor. *Educational Leadership*, 37, 15-27

Ghozali, I. (2009). *Aplikasi analisis multivariat dengan program SPSS* [Multivariate analysis application with SPSS program] (2nd ed.). Semarang: Universitas Diponegoro.

Kaplan, R.M, & Saccuzzo, D.P. (1982). *Psichological testing: Principles, application, and issues*. Monterey: Brooks/Cole.

Loder, C.P.J. (Ed.) (1990). *QA and accountability in higher education*. London: Kogan Page.

Rockwell, K., & Bennett, C. (2004). *Targeting outcomes of programs: A hierarchy for targeting outcomes and evaluating their achievement*. Faculty publications: Agricultural Leadership, Education & Communication Department. Retrieved on December 2, 2009, from http://digitalcommons.unl.edu/aglecfacpub/48/.

Santoso, S. (2014) *Seri solusi bisnis berbasis TI: Menggunakan SPSS untuk statistik multivariat*. Jakarta: Elex Media Komputindo.

Sudijono, A. (2003). *Pengantar evaluasi pendidikan* [An introduction to educational evaluation]. Jakarta: PT Raja Grafindo Persada.

Tabachnick, B.G & Fidel, L.S.(1983). *Using multivariate statistics*. New York, NY : Harper & Row.

Widoyoko, S. (2013). Pengembangan model evaluasi kualitas dan output pembelajaran IPS di SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, 11(1). doi:http://dx.doi.org/10.21831/pep.v11i1.1417