The influence of the Kirkpatrick evaluation level on the quality of learning systems in UPBJJ-UT Ternate students: Empirical studies on open university

 Effendi M [a] *Zainuddin* [b]* *MS Ahmad* [c]

[a] *English Education, Faculty of Teacher and Science Education, Universitas Terbuka, Jakarta, Indonesia*

[b] *Accounting Department, Faculty of Economics and Business, Universitas Khairun, Ternate, Indonesia*

[c] *Informatics Engineering, Faculty of Engineering, Universitas Khairun, Ternate, Indonesia*

ARTICLE INFO

ABSTRACT

This research aims to analyze the influence of the Kirkpatrick evaluation level on the quality of learning systems in UPBJJ-UT Ternate students. Kirkpatrick evaluation was conducted through four levels: reaction, learning, behavior, and results. This research is a quantitative study with a sample of UPBJJ-UT Ternate Students who registered in 2020.1 and followed tutorials in Ternate, Falabisahaya, and Bobong. The sampling technique used is purposive sampling with a sample number of 84 respondents. Data collection techniques using questionnaires. The data analysis technique used in this study is multiple linear regression analysis with smart-PLS static tools. The results showed that all attributes of Kirkpatrick's evaluation level of reaction, learning, behavior, and result levels had a positive and significant effect on the quality of learning at UPBJJ-UT Ternate open university.

Introduction

Today Education in Indonesia, in general, is faced with quality challenges. This challenge is not negotiable so that the Indonesian nation can face such fierce global competition. A qualified country will positively be correlated with the increase in Human Resources (Mukhid, 2007). Universities are indeed required to increase human resources, which is undoubtedly inseparable from efforts to improve the skin of the learning system. As one of the mega universities, Open University (UT) offers the advantages of a distance learning system (SBJJ), emphasizing high flexibility tailored to users' needs to face competitive situations with other educational institutions that provide programs with distance education mode. In addition, UT still has to defend itself concerning the ability to offer education that can create a positive and quality impression to be included in the ranks of other excellent universities in Indonesia. UT provides opportunities to prospective students who desire to study but are constrained by time, age, and college place. UT, inaugurated on September 4, 1984, has the vision to become a world-quality Open and Distance Education (PTJJ) institution in producing higher education products and implementing, developing, and disseminating PTJJ information 2021 (Djarojat, 2019)

Studying at the Open University is technology-based distance learning. Learning Services at the Open University include face-to-face learning, tuweb, online tutorials, and coursework (Anfas et al., 2018). The study aims to test the influence of Kirkpatrick evaluation level attributes on the quality of learning systems. Kirkpatrick is one of the experts in evaluating training programs in

human resource development (HR). The evaluation model developed by Kirkpatrick is known as the Kirkpatrick *Four Levels Evaluation Model.* Evaluation of the effectiveness of learning programs, according to Kirkpatrick (1998), includes four levels of assessment, namely *reaction,* learning, behavior, and *results.* The evaluation process uses a questionnaire that contains questions related to student satisfaction over the tutorial process at the reaction level. The assessment system uses performance assessment at the learning level by making an assessment rubric. At the behavior, the class is done to get an idea of student behavior towards learning activities. There is a result level more emphasized on students' final results.

Self-study is still a big problem for Open University (UT) students who carry out their learning through distance learning systems. To help UT students who have difficulty studying, UT provides learning assistance in face-to-face tutorials and online tutorials. Given the importance and strategic role of face-to-face tutorials for UT to help UT students who have learning difficulties, the quality of tutorials should constantly be improved from time to time. The best way to improve the quality of UT's face-to-face tutorial program is to evaluate the program continuously. Considering the Kirkpatrick model is one type of evaluation that can be done. Although much challenged and criticized by researchers, many researchers continued to develop their research models using Kirkpatrick's theory (Reio et al., 2017). This is because this model can be combined with other fields and avoid errors in the evaluation process (Li et al., 2008).

Literature Review

According to Ruskanda (2018)**,** education and training are activities to obtain more qualified human resources. They have a good mental attitude, honest behavior, discipline, and devotion, to improve service and protection to consumers, especially students. Therefore, the evaluation of learning is critical. They are needed to improve the quality and quality of education. In the quality management system, Zuhrawaty (2009) stated that principles used for continuous improvement, known as six (6) Principles of Quality Management, namely: Customer focus, leadership, *human resource engagement,* process approach, management system approach**,** and continuous improvement.

According to Ralph G. Lewis (1994), the higher education system presents an open system model of the core functions of student learning in higher education. The distance learning system, which is a model of the available system that is proposed at the Open University, it is also following the   Regulation of the Minister of Education and Culture number 24 of 2012 on the Implementation of Distance Education in Higher Education Article 1 Paragraph 4 states that tutorials are a form of academic learning assistance that can implement face-to-face or through the use of information and communication technology (Depdikbud, 2012).

Holmberg (2005: 104). explains that the purpose of tutorials in distance education is to: (1) Encourage student interest and motivation through contact with tutors and counselors, (2) Encourage and facilitate student learning in applying knowledge and skills to complete tasks where the task is checked, discussed with tutors and improved based on feedback and comments provided by tutors, (3) Provide opportunities for each student to develop thinking skills, (4) To measure the learning progress of each student. To support the improvement of the quality of distance learning in an open university, Dalam was conducted by evaluating learning through the implementation model Kirkpatrick to measure the quality of the learning system.

State that Kirkpatrick's evaluation model evaluates a program. There are four levels in this evaluation model: *reaction, learning, behavior, and result.* Kirkpatrick (1998:20) fattens that education is how participants change attitudes, increase knowledge, and improve skills following a program. Evaluation is a systematic process that provides information about program achievement. Its means that the evaluation includes information on whether the goal has been achieved or not.

 Evaluation is a systematic process of collecting data, news, and interpretation to be used as a basis for policymaking, decision making, or other program creation as a result of the evaluation. This can be information that can be used to revise, stop, or continue the program. Research on evaluations aimed at identifying services was conducted by Nell &Cant (2014)**,** a South African university. The results showed that students' perceptions of service quality and overall service satisfaction levels were slightly above average. Badu (2013) researched performance evaluations conducted to assess student products or projects. And Dewi & Kartowagiran (2018) investigated to evaluate the internship program by using an evaluation program by implementing the Kirkpatrick model. Based on this, research is intended to test the influence of the attribute level model Kirkpatrick on the quality of system learning at the University. Open especially UPBJJ-UT Ternate.

Research and Methods

The research method used is quantitative research. The data used are primary data obtained directly from the source. The research design used in this study is ex post *facto,* a study conducted to research events that have occurred and then look back to find out the factors that cause the event.

The data was collected using questionnaires distributed to 84 respondents, namely students who followed face-to-face tutorials at UPBJJ-UT Ternate with sample areas of Ternate, Falabisahaya, and Bobong cities. Kirkpatrick's evaluation consists of four levels of measurement: 1). Reaction Level;2). Learning Level;3). Behavior Level; and 4). Result Level.

Reaction Level means measuring customer satisfaction. Learning programs are considered adequate if the learning process is fun and satisfying for students to be interested and motivated to learn and practice. In other words, students will be motivated if the learning

process runs satisfactorily for students, which will eventually lead to a pleasant student reaction. Conversely, if students are not satisfied with the learning process, they will not be motivated to follow further learning (Partner, 2009). In the aspect of Learning Level, there are three things taught in the learning program: knowledge, attitude, and skills. Students have learned if they have experienced a change in attitude, knowledge improvement, and skills improvement (Effendi et al., 2022). Therefore, to measure the effectiveness of the learning program, then these three aspects need to be measured. Without a change in attitude, increased knowledge, or improvement of skills in students, the program can be said to fail.

Evaluating learning assessment is called the assessment of learning results (output). Therefore, in the measurement of learning, measurement means the determination of one or more of the following, namely: (1) knowledge that has been studied; (2) change in attitude; and (3) skills that have been developed or improved (Effendi et al., 2021). Evaluation at this Behavior Level is different from evaluating attitudes at level 2. Attitude assessment at level 2 evaluation is focused on attitude changes that occur when learning activities are carried out so that they are more internal. In contrast, behavioral assessments focus on behavior changes after students are in the community.

Changes in behavior happen in the community after students follow the learning program. In other words, what needs to be evaluated is whether students feel happy after participating in learning activities and returning to the community (Watkins et al., 1998). The Result Level evaluation focused on the final result of students after participating in a program. According to Kirkpatrick (2009), which included the results of a learning program, including productivity increases, improved quality, decreased costs, decreased quantity of work accidents, decreased turnover (turnover), and increase in profits. Some programs have the goal of improving work morale and building better teamwork. In other words, it is the evaluation of the program's impact (Ruskanda, 2018).

The purpose of the study was to analyze the effect of reaction level, learning level, behavior level, and result level on the quality of learning systems. The number of samples used was 84 respondents. Furthermore, data analysis will be using the Partial Least Square (SEM-PLS) technique. SEM-PLS is a multivariate method that combines factor analysis and multiple regression analysis, allows researchers to examine interrelated relationships between indicators in measuring latent constructs simultaneously, and looks at a series of connections between latent constructs (Hair. et al., 2015).

There are two types of SEM analysis: covariance-based SEM (CB-SEM) and variance-based SEM (VB-SEM). CB-SEM is also known as SEM only, while VB-SEM is also known as PLS-SEM or PLS only (Hair Jr. et al., 2017). Hair Jr. et al. (2017) explained that PLS is a variance-based SEM statistical method that simultaneously tests measurement models. In a simple sense, PLS provides precise and efficient estimation techniques for simultaneously estimating a series of equations (measurement and structural). It is characterized by two essential components in pls, namely (1) measurement model (outer model) and (2) structural model (inner model).

The outer model allows researchers to assess the contribution of each indicator in measuring construct (validity) and how reliable the measurement scale is in measuring constructs (reliability). In contrast, the inner model allows researchers to analyze the path of influence of independent constructs (exogenous) to dependent (endogenous) constructs (Hair Jr. et al., 2017)

Findings and Discussions

Findings

Convergent validity test results are known to have one indicator on the research. Suppose outer loading value smaller than 0.50, thus causing a p-value value of more than 0.50. In that case, the indicator is X1.4, so the X1.4 hand is concluded invalid in measuring the reaction level variable and does not have convergent validity. So it cannot be used for further analysis. After removing the weak indicator, the test results depicted in figure 1.

The evaluation of the inner model is from the R-Square value or coefficient of determination. In measuring the R^2 level, it will have a range of values 0 to 1. In his book, Hair Jr. et al. (2017) 0.75, moderate at 0.50, and weak at 0.25. The problem that may arise is that if there are insignificant independent variables with dependent variables that can cause the value of R^2 to increase, then Adjusted R^2 can be used to eliminate the bias. Based on data processing with PLS, R^2 adjusted value is generated by 0.871, which means that the influence of reaction level, learning level, behavior level, and result level on the quality of the learning system is 87.1%, while the rest is other variables explain 12.9%. The R^2 adjusted value of 0.912 falls into a substantial or powerful category. This indicates that the prediction of the influence of the quality of the learning system based on reaction level variables, learning level, behavior level, and result level is perfect. The subsequent evaluation of the inner model is seen from the value.

The value indicates the contribution of an independent construct to the R2 of its dependent construct. Different construct contribution levels of 0.02, 0.15, and 0.35 suggest that separate variable constructs have small, moderate, or significant contributions to conditional constructs.
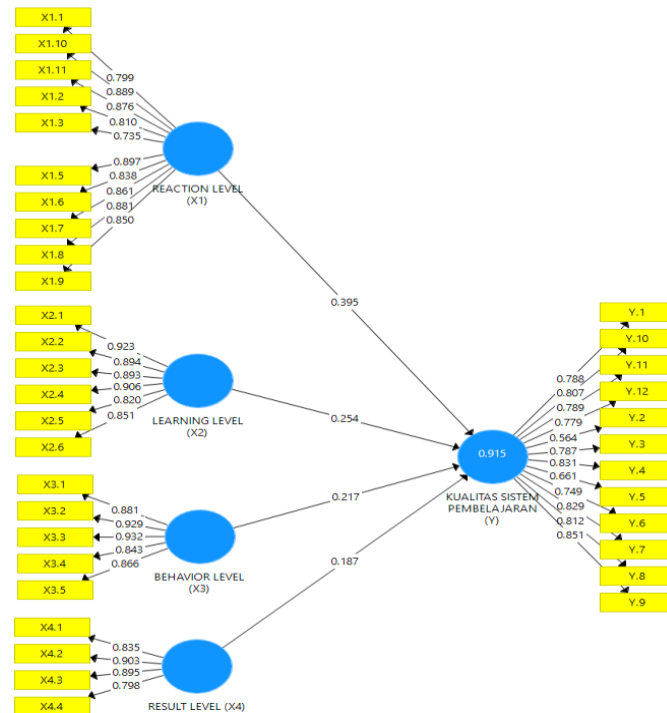
Figure 1: Model Estimate Results

Here is the result of the calculation f2 on each exogenous construct:

Table 1: Value f^2 effect size

Independent Variables	f^2	Level
Reaction Level (X1)	0.524	Big
Learning Level (X2)	0.212	Moderate
Behavior Level (X3)	0.181	Moderate
Result Level (X4)	0.172	Moderate

Source: processed data, Researcher 2022

The most extensive value f2 is at the construct reaction level (0,524); this shows that the construct that contributes the most to the change in the quality value of the learning system is the reaction level, followed by the learning level (0.212), behavior level (0.181), and the resulting level (0.172).

Predictive Relevance (Q-Square)

Measurements are tested using Q^2 blindfolding testing and a model can be said to meet predictive relevance criteria if the coefficient is higher than the value of 0 (Hair et al.,2017). In measuring levels, the relative size of predictive relevancies., values 0.02, 0.15, and 0.35, indicates that independent variable constructs have small, medium, or considerable predictive relevance for certain dependent variable constructs. Analysis results in Q^2 value of 0.494 that meets criteria more than 0 can be classified as sizeable predictive relevance, meaning variable reaction level, learning level, behavior level, and result level are very relevant in predicting the quality of learning systems.

Model Conformity Testing (Model Fit)

Evaluation of the fit model is done to see if the model used in this study is appropriate or not. In measuring the fit model, it is done with a Standardized Root Mean Square Residual (SRMR)value. SRMR represents the degree of difference between model and data, and a small weight is close to zero expected. Hair et al. (2017:208) explained the SRMR value limit of less than 0.08 indicates a fit or fit model (good fit). The results of the evaluation of the fit model on the PLS model in Appendix 6 showed an SRMR value of 0.055, this value is smaller than 0.08, so it was concluded that the model in this study had a good fit model.

Hypothesis Testing with Path Coefficients

Using PLS analysis, a path coefficient table of bootstrapping results analyses the significance of direct influence. In 2-tailed testing, the research hypothesis is acceptable if the that-statistic ≥1.96 or a p-value lower than the error rate (α) of 5%. Here are the path coefficient values (original sample estimate), t-statistics, and p values in the inner pls model:

Table 2: Analysis of The Significance of Influence Between Variables

Hip.	Influence Between Variables	Original Sample	T Stat	P Values
H_1	Reaction Level (X_1) Learning System Quality (Y)→	0.376	6.143	0.000
H_2	Learning Level (X2) Quality of Learning Systems (Y)→	0.239	4.267	0.012
H_3	Behavior Level (X3) Learning System Quality (Y)→	0.209	3.510	0.002
H_4	Result Level (X4) Learning System Quality (Y)→	0.193	3.134	0.003

Source: processed data, Researcher 2022

The results of hypothesis testing using pls bootstrapping results can be matched in Table 2 above can be explained as follows:

i.   The coefficient value of the effect of reaction level on the quality of the learning system is 0.376 (positive) with T-statistics of 6.143 and a p-value of 0.000 (smaller than α = 5%), this indicates a reaction level has a significant positive effect on the quality of the learning system, meaning that the higher the reaction level, the more quality the learning system. Based on these results, the first hypothesis states the reaction level affects the quality of the learning system acceptable (H_1 accepted).

ii. The coefficient value of the influence of learning level on the quality of the learning system is 0.239 (positive) with T-statistics of 4.267 and a p-value of 0.012 (smaller than α = 5%), this indicates that learning level also has a significant positive effect on the quality of the learning system, meaning that the higher the learning level, the more quality the learning system. Based on these results, the second hypothesis that learning story affects the quality of the learning system is also acceptable (H_2 accepted).

iii. The coefficient value of behavior level influence on the quality of the learning system is 0.209 (positive) with T-statistics of 3.510 and p-value 0.002 (smaller than α = 5%), this indicates that behavior levels also have a significant positive effect on the quality of the learning system, meaning that the higher the behavior level, the more quality the learning system. Based on these results, the third hypothesis that states behavior level affects the quality of the learning system is also acceptable (H_3 accepted).

iv. The coefficient value of the effect of the resulting level on the quality of the learning system is 0.193 (positive) with T-statistics of 3.134 and a p-value of 0.003 (smaller than α = 5%), this indicates the resulting level also has a significant positive effect on the quality of the learning system, meaning that the higher the resulting level, the more quality the learning system. Based on these results, the fourth hypothesis that the resulting level affects the learning system's quality is also acceptable (H_4 accepted).

Discussion

Effect of Reaction Level Attributes on The Quality of Learning Systems

The results showed that the reaction level attribute affects the quality of the learning system at UPBJJ-UT Ternate. Evaluation of student reaction means measuring student satisfaction of the learning process (customer satisfaction). The tutorial process is considered adequate if it is deemed fun and satisfying for students to be motivated to learn. Catalanello & Kirkpatrick (1968: 2-9) explained that evaluating participants' reactions, in this case, students, means measuring participants' satisfaction. The program is considered adequate when the learning process is fun for participants interested in and motivated to learn and practice.

Conversely, if participants are not satisfied with the training process, they will not be motivated to pursue further training. The success of the learning process is inseparable from the interest, attention, and motivation of participants in following the course of this activity. Students will learn better when they react positively to the learning environment. The results showed that students are satisfied with the learning process studied from several aspects: the material provided; facilities available; Adequate internet network, material delivery strategy by Tutor, learning media, to tutorial schedule.

Effect of Learning Level Attributes on the Quality of Learning Systems

The results showed that the learning level attributes affect the quality of learning systems at UPBJJ-UT Ternate. According to Kirkpatrick & Kirkpatrick (2008: 42), evaluation of learning outcomes can be seen in changes in attitude, improvement of knowledge, and or modification of participants' skills after completion of the tutorial. Participants have learned if they have experienced a shift in mentality, improved ability and progress, and skills. These three aspects need to be measured to measure the program's effectiveness. Without a change in attitude, increased learning, or improvement of skills in tutorial participants, the tutorial can be

said to fail. Based on the research results on UPBJJ-UT ternate students, it can be concluded that there is a change in attitude, improvement of knowledge, and improved skills after following the tutorial process. This result is in line with the theory expressed by Kirkpatrick &Kirkpatrick (2008), which states that assessment at the learning level is also called the assessment of learning results (output). Therefore, the measurement of learning outcomes must determine: (a) what knowledge has been learned; b) change in the attitude of what has been done; c) what skills have been developed or improved.

Effect of Behavior Level Attributes on The Quality of Learning Systems

According to Kirkpatrick & Kirkpatrick (2008: 53), behavioral evaluation can be done by: (1) comparing the behavior of the control group with the behavior of program participants, (2) comparing behavior before and after attending a program, (3) surveys/interviews with coaches, superiors and subordinates of program participants after returning to work. Based on the study results, it was obtained that the attributes of behavior level affect the quality of the learning system. The results showed that the margaritas of respondents, in this case, UPBJJ-UT Ternate students stated that there was a change in behavior after following the learning process at an open university. Students feel more friendly, polite, and better in the surrounding community environment.

Effect of Result Level Attributes on Learning System Quality

The results showed that the result level attributes affect the quality of the learning system at UPBJJ-UT Ternate. Evaluation of results at this level is focused on the final development that occurs because of students after participating in a program. The final results of a learning program include increased productivity, improved quality, decreased costs, decreased quantity of work accidents, decreased turnover (turnover), and increased profits. Some programs have the goal of improving work morale and building better teamwork. In other words, it is an evaluation of the program's impact.

Conclusion

Based on the test results, the Kirkpatrick Method can be used as one of the methods used to evaluate the quality of learning systems to improve the quality of higher education. This is evidenced by all levels of Kirkpatrick evaluation, namely Reaction Level, Learning Level, Behavior Level, and Result Level positively and significantly affect the quality of the learning system. The study focused on only one University with limited user data, so efforts to generalize the study results should be made with caution. In addition, the use of questionnaires as a medium to collect data only relies on students' memories of their past experiences that allow errors to occur. Data development needs to be done by increasing the number of universities in the future. In addition, comparing the evaluation model with the Kirkpatrick model can also be done.

References

Anfas, Sudarwo, R., Umasugi, M., Zainuddin, & Widokarti, J. R. (2018). The influence of learning motivation with the technology-based distance learning system. *International Journal of Engineering Research and Technology*, *11*(3), 427–437.

Dewi, L. R., & Kartowagiran, B. (2018). An evaluation of internship program by using Kirkpatrick evaluation model. *Research and Evaluation in Education*, *4*(2), 155–163. https://doi.org/10.21831/reid.v4i2.22495

Depdikbud. 2012. Peraturan Menteri Pendidikan dan Kebudayaan Nomor: 24 Tahun 2012 tentang penyelenggaraan pendidikan jarak jauh pada perguruan tinggi

Djarojat, O. (2019). Katalog Sistem Penyelenggaraan Universitas Terbuka 2019/2020. *Universitas Terbuka*.

Effendi, M., Zainuddin, Ahmad, M. S., & Sallu, Z. (2021). Quality Evaluation Of Learning Systems In Open Universities (Ut) Using Kirkpatrick Method Based On Website. *Psychology and Education Journal*, *58*, 2273–2279. http://psychologyandeducation.net/pae/index.php/pae/article/view/4232

Effendi, M., Zainuddin, Z., & Ahmad, M. S. (2022). Implementasi evaluasi model kirkpatrick terhadap kualitas sistem pembelajaran. *Jurnal EDUCATIO (Jurnal Pendidikan Indonesia)*, *8*(1), 1–8. https://doi.org/https://doi.org/10.29210/1202221160

Hair., J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2015). A primer on partial least squares structural equation modeling (PLS-SEM). *International Journal of Research & Method in Education*, *38*(2), 220–221. https://doi.org/10.1080/1743727x.2015.1005806

Hair Jr., J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, *1*(2), 107. https://doi.org/10.1504/ijmda.2017.10008574

Holmberg, B. (2005). The Evolution, Principles and Practices of Distance Education. Oldenburg : Bibliotheks-und Informationssystem der. Carl von Ossietzky Universität Oldenburg

Li, T., Yang, Y., & Liu, Z. (2008). An improved neural network algorithm and its application on enterprise strategic management performance measurement based on the Kirkpatrick model. *Proceedings - 2008 2nd International Symposium on Intelligent Information Technology Application, IITA 2008*, *1*, 861–865. https://doi.org/10.1109/IITA.2008.168

Mukhid, A. (2007). Meningkatkan Kualitas Pendidikan Melalui Sistem Pembelajaran Yang Tepat. *Tadrîs*, *2*(1), 120–133.

Nell, C. E., & Cant, M. C. (2014). Determining student perceptions regarding the most important service features and overall satisfaction with the service quality of a higher education institution. *Management - Journal of Contemporary Management Issues*, *19*(2), 63–87.

Partner, C. (2009). *Implementing the Kirkpatrick Evaluation Model Plus. Retrieved from http://www.coe.wayne.edu/eval/pdf.*

Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A Critique of Kirkpatrick's Evaluation Model. *New Horizons in Adult Education and Human Resource Development*, *29*(2), 35–53. https://doi.org/10.1002/nha3.20178

Ruskanda, L. (2018). Implementation of the Kirkpatrick Model Training Program Evaluation. *International Journal of Advanced Research*, *6*(8), 878–892. https://doi.org/10.21474/ijar01/7590

Watkins, R., Leigh, D., Foshay, R., & Kaufman, R. (1998). Kirkpatrick plus: Evaluation and continuous improvement with a community focus. *Educational Technology Research and Development*, *46*(4), 90–96. https://doi.org/10.1007/BF02299676

Zuhrawaty. (2009). *Panduan dan Kiat Sukses Menjadi Auditor ISO 9001*. Yogyakarta. Penerbit: Medpress

Publisher's Note: SSBFNET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.