

A Return on Investment Approach to Training Evaluation in Public Sectors

Juliarti Jayani

National Graduate Institute for Policy Studies, Tokyo, Japan

Yoshitaka Yamazaki

International University of Japan

September 2012

IUJ Research Institute
International University of Japan

These working papers are preliminary research documents published by the IUJ research institute. To facilitate prompt distribution, they have not been formally reviewed and edited. They are circulated in order to stimulate discussion and critical comment and may be revised. The views and interpretations expressed in these papers are those of the author(s). It is expected that the working papers will be published in some other form.

A Return on Investment Approach to Training Evaluation in Public Sectors

Authors

Juliarti Jayani^(a) and Yoshitaka Yamazaki^(b)

^(a)National Graduate Institute for Policy Studies, Tokyo, Japan

^(b)Graduate School of International Management, International University of Japan, Niigata,
Japan

Correspondence to the authors

Yoshitaka Yamazaki, Graduate School of International Management, International University
of Japan, 777 Kokusai-cho, Minami Uonuma, Niigata, 949-7277 Japan

Tel: 025-779-1106

Fax: 025-779-1187

E-mail: yyama@iuj.ac.jp

Abstract

The primary purpose of this study is to explore how to apply a return on investment (ROI) technique for evaluate training programs of public sectors according to the training evaluation paradigm proposed by Phillips. In line with his paradigm, we also examined the effect of training programs on behavioral outcomes by applying ordinary methods against knowledge and skills development and performance improvements. Due to the unique nature of public sector business process, we developed modified data-gathering instruments properly fitted to a training program of public sectors. We investigated the Intermediate Auditor Training Program by Directorate General of Taxes of Indonesia and analyzed 113 auditors who took this training program. Our analysis illustrated that the ROI method was effectively applied for the program evaluation of the public sector. In addition, we also confirmed that the other ordinary methods made differences how to assess knowledge and skills development as well as performance improvements in the public organizational context. Consequently, we have concluded that the ROI evaluation technique will be a useful tool to determine how training programs offered in public sectors are effective in view of financial benefits.

Keywords

Training evaluations, return on investment, Phillips' evaluation model, public organization, Indonesia

Introduction

Training is vital to equip employees with required skills to cope with growing market competitiveness. It has also become increasingly important for today's organizational situations where the organization undergoes business restructuring and rapid changes as to provide opportunity for employees to learn new skills needed. Berman, Bowman and West (2009) define training as the endeavor to enhance the knowledge, skills, and abilities (KSAs) of organizational members for their better performance. Along with this definition, training programs are provided at all levels of employees. For example, new employees' training programs are aimed to provide the employees with knowledge and skills about tasks, procedures, rules and regulations necessary to perform their jobs. As another, the existing employees involve training packages through which to develop their capabilities in performing the task or to cope with new changes in organizational business processes.

Although managers recognize the magnitude of training, they tend to often prioritize training lower than other needs in organizations. Berman, et al. (2009) pointed out three reasons to explain about this managerial situation. First, the urgency of the training has been usually underestimated when comparing to other immediate needs. Second, the failure to report training benefits in term of measurable financial perspectives results in less attention. Finally, managers may hesitate to spend budget on the investment for employees whose period of stay in the organization is uncertain.

In recent years, as the notion of good governance were disseminated into public organizations, public sector administrators have emphasized to build accountability into its process. Weaver and Stares (2001) argue that good governance should be judged from the quality of the decision and policies, particularly whether to produce outcomes that are broadly efficient, equitable, sustainable, and cost effective. Given the challenges, human resource

practitioners, especially those in training and development areas, are burdened with frustration to meet such demands. More accurately, they are expected to not only present excellent programs but also prove good program results, including job performance improvements, efficient use of resources, and satisfactory returns on the training costs invested (Bell & Kerr, 1987). This situation has brought them to search for a simple yet credible way to account for the training's impact. Of those expected accountabilities, the return on training investments is a most concern to human resource practitioners.

One methodology which matches this need for human resource practitioners is a methodology which was initiated in 1970s called Return on Investment (ROI) from Jack J. Phillips. According to Phillips (2003), the methodology has been developed, modified and refined to adapt to the needs in measuring training impact as well as to fit to various kinds of situation, applications and sectors. The literature of human resources management (HRM) provides detailed guidance and various case studies on how to apply ROI in the private sector. Both private and public sectors are urging to improve efficiency and accountability in their operation. Private sector increases efficiency to allow for greater profits, while public sector pursues efficiency due to budget constraints and taxpayer's pressures. Phillips and Phillips (2002) suggested that the differences between the private and public sectors have stimulated complexities in the applications of ROI process. They found that five issues emerge as uniqueness to the application of ROI in public sector: (1) the inexistence of revenue and expenditures, (2) the absence of hard data, (3) multiple constituencies must be served, (4) government provide essential services, and (5) restricted range of options to correct problems. With the existence of these challenges, developing ROI model for public sector is not as simple as adopting the currently available methodology for private sector. Until recently, however, very few studies have explored to what extent this methodology is applicable,

effective and efficient for the public sector given its constraint on providing proper data.

Accordingly, it is necessary to certify whether ROI processes are applicable to public sectors using empirical data from public sectors. Our study will handle this issue through perspectives of Phillips' model of training evaluation that includes the ROI processes as well as behavioral outcomes as the effect of training of public sectors.

Review of literature

Necessity of training evaluations

Evaluation involves comparing actual achievements toward a set of prior planned objectives, and it usually entails thorough study and careful judgments. Hamblin (1970) as cited in Sims (1993) defines training evaluation as the effort to get information as the effects of a training program and the judgment the importance of the training based on the information collected. Training evaluation can also be seen as an attempt to answer the question whether the training work and how to improve it in the future.

The magnitude of training evaluation is unquestionable. Training evaluation benefits all parties involved in the training program such as program planners, trainers, managers and trainees. Program planners and trainers obtain information as a feedback about the effectiveness of their program and seek ways to improve it whenever necessary. As for the managers, training evaluation provides information about training cost and its impact to organization. Murphy, Moller and Bencoter (1997) suggest that training evaluation justify the spending on training which link the training department with senior management. Moreover, it provides evidence about the effect of training on the change in employees' behavior and ultimate impact to organizational performance. Training participants will be able to assess their knowledge and skills via training evaluation.

Most training designs incorporate evaluation as part of its systematic process. This phenomenon shows that the concept of training evaluation has been recognized as important part of training cycle. However, literature provides different evidence on the application of evaluation in real world. Bell and Kerr (1987) claims that although the concept of evaluation has widely accepted, the actual implementation of evaluation has lagged behind. Another study conducted by Clement and Walker (1979) as cited in Bell and Kerr (1987) predicted that time spent by training professionals for planning and organizing is growing. Meanwhile, the portion of time they spent for evaluation is unpromising. Literature shows several possible answers to why training and development practitioners tend to neglect evaluation.

Fundamental reasons are provided by the following studies. According to a survey conducted by Bell and Kerr (1987) to numbers of trainers, although 90 percent of the respondents believed on the essential role of evaluation, they did not perform it because the organization where they worked were not required them to do so. Murphy, et al. (1997) showed that organizational culture plays an important role in deciding whether to conduct evaluation. They found that organizational resistance has become the obstacle in conducting evaluation. This resistance culture includes ignorance for quality, concern only to production, less knowledge about the value of evaluation, and no access to data. Sims (1993) argued evaluation by nature is difficult and time consuming which prevent people from doing it. Another reason is they simply assume that the training program work so there is no need for evaluation.

Training evaluations in public sectors

Sims (1993) found that although training evaluation is arguably the most important aspect of training cycle, it has become the least developed aspect of training process in public sector. All fundamental reasons discussed previously are applicable to public sector as well. In

some cases, training professionals often design and execute training programs without any intention to evaluate the training processes and its results. Training programs sometimes become part of routine activities which is repeated from year to year. Therefore, they often ignore the importance of evaluation. However, if the number of non-evaluated-trainings continues to grow, the problem in accountability will also arise (Bell & Kerr, 1987). Since evaluation is a way to justify cost effectiveness, training accountability should lean on reliable evaluation mechanism.

In recent years, as a notion of good governance spread into public sector organizations, the public sector administrator moves from activity driven paradigm to result based paradigm (Phillips & Phillips, 2002). In activity driven paradigm, the administrator tend to create as many programs as possible by consuming all available resources and report the result of each program based on input-focused criteria such as the number of programs, participants, hours, cost and content. Meanwhile, in result based paradigm, need analysis is performed to identify program urgency. Furthermore, a variety of methodologies to align the program objectives to business priority is also put in place. One of these methodologies which are initiated in private sector is a methodology called ROI evaluation framework.

Phillips and Phillips (2002) provides several examples on how public sector organizations are using ROI methodology to meet their challenges in an increasingly more accountable and efficient environment. Various types of organizations from small local governments to state governments, from major cities to national and federal programs applied the ROI methodology. However, although the interest to the application of ROI methodology is arose in public sector, challenges are inevitable given the unique characteristic of its business process. Phillips & Phillips (2002) claim that there are several myths which should be dispelled for public sector unit since its existence potentially hinder the implementation of

successful ROI in public sector, those myths include the following five items: nonprofit orientation, unavailability of hard data, multiple stakeholders, the essence of government service, and restricted range of options to correct problems.

The first myth, nonprofit orientation, is characterized as such that public sector organizations do not generate neither revenue nor profit which leads some people to argue that the application of ROI become not feasible. However, there are many ways to develop ROI by using numerator other than profit and revenue such as productivity, quality, time improvement as well as direct cost savings through efficiency enhancement. Unavailability of hard data is the second myth in which some people believe that only intangible and soft data is available in public sector. In contrast, there are actually hard data that can be used for ROI purposes such as output, quality, cost and time. The third myth is multiple stakeholders. In public sector, the number of stakeholder to whom the result of ROI evaluation needs to be reported is more than those in private sector because it also includes the law maker and the tax payers. This situation creates slightly more complexity in communicating the result of ROI process. The essence of government service is the fourth myth. The services provided by government are essential, regardless of the accountability. In other words, there is no such option to change or replace the program although it is not working properly. However, Phillips and Phillips (2002) states that major programs should be subjected to detailed evaluation, because even though the program itself cannot be altered, the efficiency and effectiveness of the program can be improved. Finally, the fifth myth is the restricted range of options to correct problems. In government sector, when a program has major problems, it is almost impossible to make radical changes or simply discontinue the program. Yet, many options are available to improve the program's efficiency, effectiveness and its connection to the desired result.

Phillip's training evaluation paradigm

Analysis about training effectiveness in this paper uses evaluation framework so called The Phillips Five-Level ROI Framework. This framework originated from the Kirkpatrick's four levels of evaluation by adding fifth level so called Return on Investment (ROI).

After forty years, perhaps Kirkpatrick's model of evaluation is the most common reference for training evaluation. Hale (1998) argues Kirkpatrick's model of evaluation has become a common reference for conducting analysis of training effectiveness. Bates (2004) suggests that the popularity of this model is based on three characteristics. First, Kirkpatrick's model provides an easy to understand and systematic approach to training evaluation. Second, this model allows the training professionals to account the result of their training programs in business terms. Finally, Bates argues that the four levels of evaluation become popular because of its simplicity. It is simple because it provides straight forward guidance in executing the evaluation. Additionally, this model uses post-training measurements which means less work need to be done for the purpose of evaluation before the training itself.

Although the four levels approach has received an overwhelming popularity, some studies address its limitations. Bates (2004) criticized the four-level model as an oversimplified perspective of training effectiveness because this model neglects individual or situational impacts when evaluating training. This notion implies that the model is incomplete because it fails to include the characteristic of the participants, work environment and organization into the evaluation process. Another critique is that the model simply assumes the causal linkages between each level. Positive "happy sheet" response lead to better learning result which allows for better application in work place which finally result in positive organizational performance. The debate over pros and cons on the model still takes place.

Among some limitations discussed previously, by considering the tremendous popularity of accountability trend, the inability to present training result in monetary value is the most serious flaw to Kirkpatrick's model. The model fails to provide basis for constructing costs and benefits analysis of the training result which make it impossible to account for training results in money term. Hale (1998) argues that effective evaluation models require placing a value on current contextual situations in order to examine the degree to improvement. Some researchers, recognizing the shortcomings of Kirkpatrick's four-level approach, have attempted to modify and add to this basic framework.

One of the latter developments in evaluation methodology is an approach called, "The Phillips Five-Level ROI Framework." Originally, ROI is a measure of corporations' profitability to assess how effectively the company uses its capital to generate profit. In other words, ROI compares total benefit of a program to its total cost. This concept later is used by a wide range of disciplines which intends to seek the measurement using cost and benefit analysis. In term of training, ROI is a measure of the monetary benefits obtained by an organization over a specified time period in return for a given investment in a training program. Looking at it another way, ROI is the extent to which the benefits (outputs) of training exceed the costs (inputs).

Phillips (2003) provides a simple and logical model for calculating return on investment in training programs. This model consists of step-by-step approach so that the process is manageable which makes it possible to tackle one issue at a time. As described earlier, the Phillips framework consists of five levels of evaluations: The first level describes the evaluation of reactions, satisfactions, and planned actions; the second level represents learning; the third level involves application and implementation; the fourth level indicates

business impact; and the fifth level shows return on investment. Each of the five levels is illustrated as follows:

Level 1: Reaction, Satisfaction, Planned Action

Measures how participants' satisfaction about the training, as well as their plan to apply what they have learned. Phillips (2003) suggests that almost organization perform evaluation at this level. However, although this level of evaluation is important as a customer satisfaction measure, a favorable reaction does not ensure that participants have learned new skills or knowledge.

Level 2: Learning

Focuses on assessing to what degree participants absorb the materials that are provided during the training program, using tests, skills practices, role plays, simulations, group evaluation, and other assessment tools. This assessment is necessary to ensure the participants comprehend the materials and know how to use it properly. Yet, a good result at this level does not automatically mean that what is learned will be applied on the work place.

Level 3: Application and Implementation

Measures how training program changes the participants' behavior at workplace through the application of what they have learnt from the program. A variety of follow up methods are used to determine whether participants applied what they learned on the job. The frequency and use of skills are important measures at this level. While Level 3 of evaluation is important to gauge the success of the application of a program, it still does not guarantee there will be a positive business impact in the organization.

Level 4: Business Impact

Measures focuses on the actual results achieved by program participants as they successfully applied what they have learned. Typically, include output, quality, costs, time,

and customer satisfaction. Although the program may produce a measurable business impact, there is still a concern that the program may cost too much

Level 5: Return on Investment

The ultimate level of evaluation compares the monetary benefits of the program with the program costs

While almost all training organizations conduct evaluations to measure satisfaction, very few conduct evaluations at the ROI level (Phillips, 2003). In most situations, training organizations hesitate to perform the ROI calculation because of the perceived difficulties and complexities of the methodology. Moreover, to end up with ROI calculation, it is also necessary to evaluate the other levels which make the process become very expensive. Due to chain impact through all levels as assumed in Kirkpatrick's four-level model, it is difficult to conclude that the results achieved were actually caused by the program if the evaluation is not conducted comprehensively. Therefore, it is recommended that evaluation be conducted at all levels when a ROI level of evaluation is planned. Based on a framework by Phillips, the following are five steps to be conducted in calculating ROI: (1) data collection; (2) isolating data on the effects of the training; (3) assigning monetary value to isolated effects; (4) calculating the total course cost; (5) calculating the return on investment.

Data collection

Phillips (2003) defines six categories of data that can be used in measuring ROI: Organizational Performance Records, Participants, Manager of Participants, Subordinates of Participants, Team/Peer Group, and Internal/External Group. Among these sources of data, performance records are the most reliable tools, followed by Manager of Participants. However, he continues to say that if the data collection is hindered by limitation of the existence of performance records, data that are collected from the participants will also

provide valuable information. A similar idea is provided by Fusch (2001) shows that the number of participants of a training program may be small but enough to provide information necessary to evaluate the program through the return-on-investment model applying the questionnaires of participants and those of the supervisor post-training. Additionally, Bartel (2000) in his study concludes that to calculate ROI, companies should use internal data as much as possible because it provides more control to all factors other than training which may have potential impact on ROI. In terms of how to collect the data, Philip (2003) suggests several ways such as surveys, questionnaires, interviews, focus groups, tests, observation, and performance records.

Isolating data on the effect of training

After training programs, there is usually an improvement on employees' performance. Yet, how much of this improvement was caused by training program is not clear until a certain mechanism is taken to measure the causal relationship between training and performance. This mechanism in ROI model is named as isolating the effect of training. Basically, steps are performed to explore and determine the amount of performance directly related to the program. Phillips (2003) argues that a robust ROI approach should be able to address which improvements are directly related to the training. He presents several techniques that have been used by organizations to cope with this issue, including a control group arrangement; trend lines; a forecasting model; participant estimates; supervisors of participant estimate; senior management estimates; experts estimates; and customer inputs.

Assigning monetary value to isolated effects

To be able to measure ROI, data from Level 4 of evaluation are converted into monetary values. Various techniques are available for this purpose; however, the selection of techniques to be used heavily depends on the types of data used in the model.

Calculating the total course cost

ROI process compares benefits of training to its total cost. Therefore, tabulating the cost of training is vital part in calculating the ROI. The costs of training should include all possible cost components related to the training, ranging from design and develop the program, training delivery cost as well as evaluation cost.

Calculating the return on investment

After the monetary data for benefits and costs can be matriculated, ROI then are calculated using the formula as in Figure 3. Interpreting ROI result is more or less the same as those of ROI application in financial sector. For example, if after calculation we get ROI of training investment equals 50%, it means that after the costs are recovered, an additional 50% of the training are reported as “earnings”. However, not every ROI value are positive. In the case that training investment is having negative ROI value, it means that although the training is effective to achieve certain set of objectives, the costs of the training exceed its benefits. It is also to say that the investment is too expensive compare to its benefits. The formula for calculating ROI of Fusch (2001) as described below:

$$ROI = \frac{\text{Training Program Benefits}}{\text{Training Program Costs}}$$

$$ROI (\%) = \frac{\text{Net Training Program Benefits}}{\text{Training Program Costs}} \times 100$$

Method*Research sites*

For this research, we were allowed to access Indonesia government employees who work for Directorate General of Taxes (DGT) that is a vertical unit under the Ministry of Finance whose function is to formulate and execute policies and technical standardization in

taxation area. With 33,000 employees, DGT plays a very important role since it generates major revenue to support government spending. Currently, tax revenues are accounted for 80 percent of total government income.

As the call for more efficient and sound public sector management has risen immensely, the Directorate General of Taxes (DGT) under the Ministry of Finance of the Republic of Indonesia has also taken series of reforms to apply the idea of good governance. Firstly, in the year of 2002-2007, the reform so called bureaucratic reform was undergone. This first reformation package concerns mostly about changing the attitude of employee particularly on corruption and applying code of conduct. During the implementation stage, many problems occurred due to lack of employees understanding on what they are expected to be, others happened because of the absence of strong and reliable enforcing mechanism. As a result, the reformation process seemed to be forced.

The next reformation package is known as Program on Indonesian Tax Administration Reform (PINTAR) Project. As stated in the project report, this project aims to increase taxpayer compliance by increasing the efficiency and effectiveness of the DGT and to improve good governance in tax administration by strengthening transparency and accountability mechanisms. A higher level of voluntary compliance combined with increased integrity of tax administration due to better transparency and accountability mechanisms should result in a reduction of the tax gap and in addition contribute to an improvement of the investment climate in Indonesia. The strategic portion of the PINTAR Program is composed four major components. Component A includes Core Tax Administration Enhancement (composed of Registration, Returns Processing, Document Management, Accounts and Infrastructure enhancements). Component B represents Human Resources Management (personnel, payroll and training enhancements). Component C relates to Compliance Enhancements (composed of

Audit and Collections automation). Finally, Component D involves Oversight and improved systems management.

Among other areas in PINTAR project, the one of the authors works closely in the Component B, the transformation of Human Resources Management (HRM). This component focuses on the improvements to support the reformed tax administration processes under PINTAR through reformed HRM policies, improved training and enhanced training capacity and strengthened accountability, integrity and transparency. Significant changes to HRM policy is put forward to address the way in which the DGT civil service population is recruited, retained, managed and developed and the way in which capacity is built and sustained and accountability will be enforced and integrity and transparency enhanced.

Currently, all PINTAR features are being developed. In fact, PINTAR has allowed training and development division of DGT to do more about their training programs. The development of assessment centre, competency based training, on the job training program and e-Learning are among the new development in training and development area. At the same time, the need to improve personnel capacity through training and development activities to support on-going changes in tax policies and regulation are unavoidable. As a result, the budget allocates to training and development is huge and continues to grow every year. However, the rapid growth in budget is not supported by proper accountability measures. Until recently, training and development division has made no attempts to identify the training contribution to the organizations in terms of neither performance improvement nor monetary unit. In most of its training programs, evaluation was undergone by using the participants' reactions sheet and skills tests. Moreover, the implementation of balance score card as organizational performance measurement as part of PINTAR project, the need to be accountable for this budget has emerged significantly. Therefore, learning and development professionals are finding it

increasingly necessary to show the monetary value of the training programs they are bringing to the organization.

Targeted training programs to be evaluated

The object of ROI calculation in this study is “Intermediate Tax Auditor Training Program” of Directorate General of Taxes of Indonesia for the year of 2011. This training program is provided for tax auditor to update their knowledge and skills in the area of tax regulation, tax auditing techniques and several soft skills. In the year of 2011 this training was conducted in five batches with total participants of 427 tax auditors. The tax auditor training programs are divided into three categories: basic, intermediate and advance level in which every level has its own requirement and objectives. The measurement of the ROI requires the training participants to be able to compare their performance before and after the training program. Given this requirement, the intermediate level of auditor training is qualified to be selected as the ROI calculation object because the participants of this level are already tax auditors before their participation in the training. Meanwhile, the basic level is provided for tax auditor candidates and the advance level was not conducted during 2011.

Sampling procedures

The research methodology is modeled after the Kirkpatrick’s four levels of evaluation and by adding the fifth level from Jack J. Phillips so called Return on Investment (ROI). To narrow down the research scope, this study focuses on the 3rd and 4th level of evaluation from Kirkpatrick and measuring the ROI using Phillips concept. Level 3 is aimed to see whether the participants are having opportunity and/or able to apply what they have learned in the training program at work place. Meanwhile, Level 4 of evaluation is aimed to measure how the training has improved things on the jobs. Level 5 or Philips ROI model is aimed to cost-benefit analysis of the training. The combination of these two models allows us

to answer the questions whether the training program is effective and whether it is too costly to conduct the training.

Phillips (2003) argues that performance records are the most reliable data for ROI calculation. However, in the case of this research context, current performance records are not eligible to be used as the basis for ROI measurement. First of all, tax auditors do not have authority to set their job target. The numbers of job they need to accomplish are assigned by the supervisor based on fair jobs distribution among employee. It is to say that, although the tax auditor are equipped with better knowledge through training programs, it does not necessarily mean they achieve more in terms of quantitative performance measure. However, literature provides numerous case studies and alternative solutions for similar situation. Learnt from the studies by Fusch (2001) and Atkins (2002), the data will mainly be collected by means of questionnaire and supported by internal data from DGT.

The questionnaire was first sent for trial by email to some respondents to assure that they can receive the file and they can understand the questions given in it. Unfortunately, some technical obstacles occurred. For example, some respondents failed to receive the questionnaire due to internet security system in the DGT, and some other, although they managed to receive the questionnaire, they cannot open it because of the absence of necessary software. Coping with these obstacles, several changes were made such as converting the questionnaire file type and providing more alternatives file type to allow greater flexibility for the respondents.

On second week of April 2012, the questionnaires were sent via email to all 427 respondents who participated in Intermediate Auditor Training Program 2011. Three weeks later a reminder email was sent to those who have not replied the email. After 8 weeks of data gathering process, from second week of April until second week of June 2012, 113 filled-questionnaires are received. The respond rate is about 26.5 percent of total questionnaire

sent. At least there are three possible reasons for not responded questionnaire, the first and the most common reason for un-responded questionnaire is that the tax auditors are very busy with their workload which makes it difficult to ask them to fill the questionnaire. Secondly, some of the respondents are out of reach because they seldom check their email and there is no other way to contact them except via email. Finally, the respondents do not reply the questionnaires due to culture where people often ignore any kind of questionnaire given to them as a result of their perception that there is no benefit for them to do so.

Instruments

As discussed previously, to end up with ROI calculation, it is necessary to evaluate at every level of Kirkpatrick's evaluation model. However, to reduce complexities and to narrow down the scope of this study, the data collection instrument which is used in this study is designed to gather data necessary for evaluation at Levels 3, 4 and 5. At Level 3, training evaluation is aimed to see whether the participants are having opportunity and/or able to apply what they have learned in the training program to the work place. Meanwhile, Level 4 of evaluation is aimed to measure how the training has improved things on the jobs. Level 5 (e.i., Phillips ROI model) is aimed to perform the cost-benefit analysis of the training.

Data gathering instruments

Phillips (2003) argues that performance records are the most reliable data for ROI calculation. However, in the case of tax auditor of DGT, current performance records are not eligible to be used as the basis for ROI measurement. First of all, tax auditors do not have authority to set their job target. The numbers of job they need to be accomplished are assigned by the supervisor based on 'fair' jobs distribution among employees. It is to say that, although the tax auditors are equipped with better knowledge through training programs, it does not necessary mean they achieve more in terms of quantitative performance measure. Yet, Phillips

offers an easily implemented method to isolate the impact of training that is by obtaining information from the participants directly. This method is contingent with the belief that participants can judge the effect of training program on performance improvement by themselves (Phillips, 2003). Because the participant is the center of performance improvements, they may have very accurate input about performance improvements. Therefore, the data for the study is mainly collected by means of questionnaire and supported by internal data from DGT.

The development of the questionnaire is based on three previous studies conducted by Fusch (2001), Harrel (2001) and Atkins (2002). Fusch (2001) in his study customized the Phillips ROI model to be used as a template that can easily be adapted as needed by Georgia Pacific Company. Harrel (2001) presented a list of questionnaire content possibilities for capturing follow up data for ROI measurement for a program on leadership development. Atkins (2002) conducted a pilot study in training ROI evaluation implemented to New York State Governor's Office of Employee Relations and used questionnaire to gather information about performance improvements. Although these three studies examined a different type of training program, they all used a similar type of questionnaire to capture participants' estimate about post training performance improvements as a basis for ROI calculation. Modification and adjustment has been made during the questionnaire development process as to fit the characteristic of Intermediate Tax Auditor Training Program.

The questionnaire consists of two parts, part A and part B, with a combination of open ended and closed ended type of questions. Part A of the questionnaire aims to access to what extent the participants are able to apply the knowledge and skills they have learned in the real work place. There are five questions in this part. Question numbers 1 and 2 basically are designed to clarify whether the participants make improvements in how they perform the

job given the materials from the training program. These questions cover hard skills as well as soft skills and are originated from the training objectives. Meanwhile, question numbers 3 to 5 are intended to seek reasons for inability to apply and participants' input for program improvements in the future.

On the other hand, part B is designed to capture data for Level 4 of evaluation that is to examine to what extent has things improved because of the training. Phillips (2003) provides various measurements as indicators of organizational improvements due to training programs. After thorough examination of the characteristics of tax auditor tasks and duties, this study uses three dimensions as the measurements of performance improvements. The first dimension describes increase in performance, measuring the increase in participants' overall performance after the training based on their perception. The participants were asked how better they do their job by comparing before and after the training accomplishment. The second dimension represents reduction in time to complete the audit (time save), measuring how faster can the participants completed their tasks. This measurement lies on the assumptions that when employees are able to perform tasks in shorter time, then the time saved can be used for other productive activities. Finally, the third dimension involves increased in the quality of output (reduction in errors). The quality of the audit report is measured in terms of the number of reduction in errors. The cost of repeating an error is the time needed to repair the task. Therefore, the reduction in errors will lead to shorter time in completing the job.

In addition to the above measures, the questionnaire also seeks information about the increase in the number of output (tax audit report). Although this performance criteria is not eligible as an indicator as discussed earlier, however, to some extent, where previously the performance of tax auditors is below the targeted level due to lack of competencies, the

information about the increase in output will actually strengthen the overall result of evaluation process.

In order to enrich the result of overall evaluation process, this study also measures one possible intangible benefit resulting from the training program. Phillips (2003) states that the intangible benefits of training are the benefits that are directly linked to the training but cannot be converted to monetary values. Although these benefits do not have monetary values, they are still very important in the evaluation process and therefore should be also disclosed as the benefit of the training. For this study, the intangible benefit to be included is employees' self confidence in conducting daily task as tax auditor.

The questionnaire was first developed in English and then translated into Indonesian. In order to ensure the translated version conveys the same meaning as the English version, back-translation process was put in place. This translation procedure is proposed in cross-natured studies by Brislin, Lonner, and Thorndike (1973). The Indonesian version was sent to all 427 participants of 2011's Intermediate Tax Auditor Training Program who spread all over Indonesia. The respondents are provided with three formats of questionnaire; (1) on line survey, (2) pdf format and (3) Microsoft Word document. The purpose is to enable respondents to choose questionnaire format that suit their situation. For those who have internet connection, they can fill in the on-line version which perhaps the greatest simplicity. Meanwhile, for those who are not able to connect to internet, can choose the off-line format which is available in pdf and Microsoft Word format.

Isolation evaluation techniques

Performance result after training would be influenced by many factors other than the training itself. Various approaches are available to isolate the effect of training. Phillips (2003) mentions that an easily implemented method to isolate the impact of the training is to obtain

information directly from the training participants. The effectiveness of this approach rests on the assumption that participants are capable of determining or estimating how much of performance improvement is related to the training program. Although some human resource practitioners argue on the reliability and accuracy of this estimate due to incapability of participants to provide such information, the advantages seems to outweigh the disadvantages. The advantages of participants' estimates are in its simplicity, easily understood, inexpensive yet originated from very credible source, the participants who produce the change or improvements.

In part B of the questionnaire, a series of questions to address three performance improvements dimensions are designed to focus on isolating the effect of training. In order to increase the credibility of the information, the respondents are asked about their confidence level on his/her answer on performance improvement. The purpose of this confidence level question is to adjust the level of performance improvements to the potential errors that the respondents might make. For example, if after the training, the respondent stated 80% of improvements but with only 70% level of confidence, then the adjusted performance improvement should be $80\% \times 70\%$ that is 56%. This adjusted improvement level then is converted to monetary values. This adjustment method using a participant's confidence level is consistent with Phillips' guiding principle number 7 for isolating the effect of the training.

Data conversion techniques

Assigning monetary values to the data is the ultimate level of the five-level evaluation framework. The data gathered from part B of questionnaire are meaningful to ROI calculation only after it is converted into monetary values. The impact of the training is calculated against participants' salaries. The first question in part B asked respondents about the percentage use of skills they have learned from the training program at their work. This estimated portion is

used as a basis for calculating the percentage of salary spent on the tasks using the acquired skills. The resulting figure is then multiplied by three components of productivity gained from three dimensions of performance improvements. Meanwhile, the productivity gain from performance improvements is calculated as described in Table 1. Three performance improvement measures are converted into monetary value. The performance improvement in term of job attitude is considered as an intangible benefit of the training and therefore is not converted into monetary term. The three performance improvement measures which converted into monetary value include increase in performance, reduction in audit completion time, and increase in output quality. Increase in performance describes the most common measure for training effectiveness that is a degree of increase in employees' performance. In this study, the increase in performance is measured from participants' perception in their overall performance. Each one percentage increase in performance is associated with one percentage gain in salary. Second, reduction in audit completion time concerns faster time in completing the audit that means employees have more time to do other productive activities. Therefore, organization gain benefit from the increase in productivity. Third, the increase in quality of audit is measured through the percentage of reduction in errors. Errors in conducting audit costs more time to redo part or/and whole of the audit. In other words, reduction in errors means a saving in time to complete the audit. Thus, every percentage of reduction in errors leads to a percentage increase in time saving. Assuming that a time saving is used in productive way, a percentage increased in time saving is associated with benefit in salary. Employees are doing more work given the same salary.

Insert Table 1 about here

Cost accumulation techniques

To be able to end up with ROI figure, the total cost of training is required. Philips (2003) provides guidance on major cost categories to be included in cost tabulation. In the case of Intermediate Tax Auditor Training Program, the expense of the training is shared between two institutions, DGT and Center for Taxation Training and Development. The latter institution is a vertical unit under Ministry of Finance which responsible for the execution of training and development in taxation area. However, for the purpose of calculating ROI for DGT, only those expenses incurred by DGT are included as the program cost. After thorough analysis and some adjustment due to the availability of data, cost components are included in this study as shown in Table 2.

 Insert Table 2 about here

Results and discussion

After 8 weeks of data gathering process as described in previous section, given the data from 113 respondents, the analysis then were conducted for three levels of evaluations, Levels 3, 4 and 5.

Level 3 evaluations

Level 3 of evaluation measure the application of knowledge and skills given in the training to the work place. Questionnaire part A accesses how things change in work place after the training. This part captures two major types of information, the improvements in term of hard skill and soft skills. In order to perform statistical analysis, a number is assigned to each scale of Likert interval that is used in the questionnaire. Value of 0 is given to “no change” and 4 to “very significant changes”. Table 3 presents descriptive statistics of the questionnaire result. Meanwhile, Figure 5 shows the distribution of respondents’ feedback about how things change after the training. Using these two figures, we can explain how

effective the training is. In terms of hard skills, with the mean of 22.5, most of the respondents' answers are located to the right of the mean. It also means that most of the participants feel a moderate to very significant changes in how they perform their jobs after the training.

However, in terms of soft skills, the distribution of answers peak around the mean. Therefore we can conclude that although there is improvement in the area of soft skills, the

improvement in hard skills is bigger. However, when we analyzed hard skills and soft skills together, 108 out of 113 or 95 percent of respondents agree that the training has improved their overall knowledge and skills ranging from moderate to very significant changes.

Therefore, we can conclude that the training is effective because it generates improvements in the way people carry out their daily job.

 Insert Table 3 and Figures 1(a), (b) and (c) about here

Approximately 5 % of respondents provide below average feedback. Among various reasons, there are two major reasons to their argumentation. On one hand, some respondents said that the characteristics of tax payers and tax cases they deal with are in general very simple, in which the knowledge and skills from the training is somehow way too advance and inapplicable to their situation. On the other hand, some argued that they have had previous knowledge that they got from other sources outside of the training, by which they feel that training does not give them any added value in terms of knowledge or skills. This finding perhaps shall be an input for the program designer to improve the criteria in determining training participants. To enhance the effectiveness of the training program, it should be given to those employees who really need the knowledge and skills offered in the program.

Level 4 evaluations

Information necessary for Level 4 evaluation is gathered using Part B of the questionnaire. Question 1 of the questionnaire asking how much the participants use the knowledge and skills they got from the training. The resulting number is to be used for calculating portion of salary for job they spent using skills they have got from the training. Figure 6 shows the respondents feedback to question 1. Approximately 80% of the respondents state that their job uses 50 up to 100 percent of the skills given in the training. The rest mention that there are too much knowledge and skills given from the training while they have limited time to apply all of them at once. Moreover, as previously discussed, due to unique characteristic of tax cases and tax payers, some training participants cannot apply the given knowledge because that knowledge becomes irrelevant.

Question 2, which originally is designed to capture the performance improvements for those who before the training are underperformed, provides unexpected result to this study. As shown in Figure 7, 85% of the respondents experience the increase in the number of audit report as the result of the training. However, since this measure generates complexity in assigning appropriate monetary value to it, it is excluded from ROI calculation. One reasonable explanation is that various uncontrolled factors may have influenced the increase in the number of output such as the increase in organizational target of audit, the help from supervisor, better team work, easier tax cases or previous knowledge other than those taught in the training. Nevertheless, this finding has bring positive image to the effectiveness of the training program

 Insert Figures 2 & 3 about here

In terms of four criteria of performance improvements as in question 3, none of the 113 respondents report zero improvements to any of those criteria. After adjusted to

respondents' confidence level, the results vary from 5% to 100% improvements in all performance criteria. The complete result of performance improvements is as shown in Table 4. This means that the training has a positive impact to performance measures although the degree of improvements may *vary* from person to person.

 Insert Table 4 about here

Level 5 evaluations

Cost summary

All available costs related to the training program are tabulated and analyzed. Cost from DGT includes participants' transportation cost, participants' salaries and benefit during the 10 days training. Because there is no record about design and development cost, it is omitted from cost calculation.

The intermediate Tax Auditor Training Program is conducted in cooperation between DGT and Training Center (an autonomous body outside DGT). DGT choose and send the participants to training center for the program, but the training is delivered by Training Center. As a result, there is a cost which incurred by Training Center. Although DGT that does not bear this cost, and after considering the amount that is 18% of the total cost, in order to fairly present all costs related to the program, this type of cost is included in the costs tabulation. The training delivery cost is as in Table 5.

 Insert Table 5 about here

Benefit summary

In order to have a monetary value for each performance improvements in Level 4 evaluation, the improvements are assigned monetary value based on monetary conversion

method in Table 1. One out of four performance improvements indicator is not converted to monetary term and is treated as intangible benefit, which is self-confidence. Meanwhile, Table 6 summarizes the cost and the benefit of training program.

 Insert Table 6 about here

ROI calculation

From the detailed analysis of cost and calculated benefit of the training program, an ROI calculation is carry out as the follow:

$$\begin{aligned}
 \text{ROI (\%)} &= \frac{\text{Net Benefit}}{\text{Total Cost}} \\
 \text{ROI (\%)} &= \frac{13,315,721,455.2 - 4,721,582,235}{13,315,721,455} \times 100 \\
 \text{ROI (\%)} &= \frac{8,594,139,220.20}{13,315,721,455.20} \times 100 \\
 \mathbf{\text{ROI (\%)}} &= \mathbf{65\%}
 \end{aligned}$$

The program benefit was Rp. 13,315,721,455.20 and the program cost was Rp. 4,721,582,235.00. The net program benefit of Rp. 8,594,139,220.20 is divided by the program cost and multiplied by 100 yielding an ROI of 65 percent. The ROI of 65 percent means that the costs are recovered and an additional 65% of the costs are reported as revenue/earnings, that is Rp. 8,655,218,945.88

Intangible benefits

The improved employees’ self-confidence is the first intangible benefit of the training program. As seen in Table 4, the range of improvement is between 10 to 100 percent. None of the 113 respondents claims zero improvement in their self-confidence. As most people notice that self-confidence is one of the important ingredients of outstanding performances, the

increase in the level of self-confidence actually contribute to goal achievements. Another intangible benefit which is found during the data collection is networking. During the training, participants interacting each other and create a communication network. When they return back to their post, this network continues and become a means of transfer of knowledge and experiences between the auditors. If one auditor faces difficulties in solving certain case, he or she can seek an example from previous solved case through this network.

Brief summary of the program effectiveness

Having discussed the findings in previous sections, this study indicated that The Intermediate Auditor Training Program has achieved its objectives by using Phillip's training evaluation model. The training program has resulted in ROI of 65 percent which also means that after all the costs are recovered, 65 percent of the cost is realized as the benefit. Therefore, the training has been conducted in an efficient manner.

Practical implications

According to the results from this study, we propose three practical implications with regard to ROI of public sectors.

Preparation for usage of ROI

When public sectors decide to perform ROI methodology, they should plan ahead with an entire training program. The most difficult part for ROI calculation may collect adequately data required for the RIO due to various data need and large population of respondents. It is far easier to collect the data when ROI evaluation and data collection methods have been considered during training program design and development. It is to say that incorporating ROI process in training cycle is necessary to enhance ease of data collection, data accuracy and result credibility.

Managerial supports

It is critical to have support from mid to top level management to ensure the success of ROI implementation. Without their support, ROI implementation is potentially not run smoothly due to resistance from many parties. For example, participants basically dislike to be asked to fill the questionnaire no matter how easy the questionnaires are. Having back up from mid to top management will actually force the respondents to take the process seriously.

Intangible benefits

A variety of intangible benefits actually reflect the success of the training program, which was self-confidence in this study. Much past confidence research indicated that employees confidence is considerably critical for organizational effectiveness and success (Luthans, Luthans, and Luthans, 2004), strong job performance (Stajkovic & Luthans, 1998), job satisfaction (Luthans, Zhu, & Avolio, 2006), and organizational commitment (Werbel, Landau, & DeCarlo, 1996). Although these benefits may not view as valuable as monetary benefits directly, they nevertheless are an important part of overall evaluation. Therefore, during data gathering process, intangible benefits should be identified, measured and analyzed.

Conclusion

In this study we examined application of a return on investment methods for public sectors using Phillips' training evaluation paradigm. Because the unique nature of public sector business process, we developed modified data-gathering instruments appropriately matched with training program of public sectors. In this study, the Intermediate Auditor Training Program by Directorate General of Taxes of Indonesia was used for analysis. Our analysis showed that the ROI method was an effective method applied for the program evaluation of the public sector, while the other ordinary methods in terms of behavioral outcomes was also beneficial for the training evaluation about knowledge and skills

development and performance improvements in public sectors. Accordingly we have concluded that the ROI evaluation method will become an effective technique process for training programs of public sectors are effective in view of financial benefits.

Finally, this research used the Intermediate Auditor Training Program by Directorate General of Taxes of Indonesia for analysis. However, in order to make more generalization about ROI processes of public sectors, promising studies should be conducted other training programs of other agencies, and other countries. It is hoped that our study provides stimulate researchers to conduct more examination of the ROI methods in contexts of public sectors.

References

- Bartel, A. P. (2000). Measuring the employer's return on investments in training: Evidence from the literature. *Industrial Relations*, 39 (3), 502–525..
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 27 (3), 341–347.
- Bell, J.D., & Kerr, D.L. (1987). Measuring training result: Key to managerial commitment. *Training and Development Journal*, 41, 53–55
- Berman E., Bowman, J., & West, J. (2010). *Human resource management in public sector* (3rd ed.). Los Angeles, CA: Sage Publications.
- Brislin, R., Lonner, W. & Thorndike, R. (1973). *Cross-cultural research methods*. New York: John Wiley and Sons.
- Condrey S.E. (2010). *Handbook of human resource management in government* (3rd ed.). San Fransisco, CA: Jossey Bass
- Directorate General of Taxes (2008). Analysis document external version 1a: in Support of PINTAR project.
- Fusch, G. E. (2001). What happens when the ROI model does not fit?. *Performance Improvement Quarterly*, 14, 60–76. doi: 10.1111/j.1937-8327.2001.tb00230.x,
- Hale, J. (1998). Evaluation: It's time to go beyond levels, 1, 2, 3, and 4. *Performance Improvement*, 37, 30–34. doi: 10.1002/pfi.4140370209.
- Harrell, K. D. (2001). Level III training evaluation. Considerations for today's organizations. *Performance Improvement*, 40, 26–29. doi: 10.1002/pfi.4140400508
- Kent, & Stares. (2001). Guidance for governance: An overview. In Kent, W.R., & Stares, P.B, (Eds.), *Guidance for Governance: Comparing alternative Sources of Public policy Advice* (pp. 1–30). Tokyo: Japan Center for International Exchange.

- Luthans, F., Luthans, K.W., & Luthans, B.C. (2004). Positive psychological capitals: beyond human and social capital. *Business Horizon*, 47(1), 45–50.
- Luthans, F., Zhu, W., & Avolio, B.J. (2006). The impact of efficacy on work attitudes across cultures. *Journal of World Business*, 41, 121–132.
- Moller, L. & Mallin, P. (1996). Evaluation practices of instructional designers and organizational supports and barriers. *Performance Improvement Quarterly*, 9, 82–92.
doi: 10.1111/j.1937-8327.1996.tb00740.x.
- Phillips J.J. (1997). *Handbook of training evaluation and measurement methods* (3rd ed.). Houston, TX: Gulf Publishing Company.
- Phillips J.J. (2003). *Return on investment in training and performance improvement programs* (2nd ed.). Butterworth Heinemann.
- Phillips P. P., & Phillips J.J. (2002). *Measuring ROI in public sector*. American Society for Training and Development (ASTD).
- Robrer-Murphy, L., Moller, L. & Benscoter, B. (1997). A performance technology approach to improving evaluation. *Performance Improvement*, 36: 10–15.
doi: 10.1002/pfi.4140360804.
- Sims, R. R., (1993). Evaluating public sector training programs. *Public Personnel Management.*, 22 (4), 591–615.
- Stajkovic, A.D., & Luthans, L. (1998b). Social cognitive theory and self-efficacy: going beyond traditional motivational and behavioral approaches. *Organizational Dynamics*, 26(4), 62–74.
- Webel, J., Landau, J., & DeCarlo, T.E. (1996). The relationship of pre-entry variables to early employment organizational commitment. *Journal of Personal Selling and Sales Management*, 16, 25–36.

Table 1. Basis for Monetary Conversion

No.	Performance Improvements Dimensions	Measurement (*adjusted with potential errors)	Basis for Value
1	Job Performance	Participants' perceived increase in overall performance	Value of 1% increase in productivity = 1% gain in salary
2	Time	Reduction of time in performing the audit (time saving)*	Value of 1 % saving in time = 1% gain in salary
3	Quality	Reduction on errors (less errors means faster time of completion)*	Value of 1 % reduce in errors = 1% saving in time = 1% gain in salary

Table 2. Cost Accumulation.

No.	Cost Item	Method of allocation
1	Salaries/Benefit of facilitator	Expense
2	Salaries/Benefit participants	Expense
3	Program materials and fees	Expense
4	Lodging and meals	Expense
5	Participants travel cost	Expense
6	Participants travel time	Expense
7	Overhead	Expense

Table 3. Brief Descriptive Statistics for Level 3 of Evaluation ($N = 113$)

Variable	Mean	SD.
No	57.00	32.76
Hardskills	22.55	5.79
Softskills	10.45	2.99
Total	33.00	8.15

Table 4. Summary of Improvements by Performance Criteria

	Job Performance	Time of Completion	Reduction in Errors	Self Confidence
Lowest	5%	5%	6%	10%
Highest	100%	90%	90%	100%

Note. The value is adjusted to respondents' confidence level

Table 5. Cost of Training Delivery

No	Cost Item	Batch I	Batch II	Batch III	Batch IV	Batch V	Grand Total
1	Material	25,531,105	25,115,680	20,970,890	26,332,940	10,395,000	108,345,615
2	Meals	79,561,500	84,466,500	80,231,750	82,280,000	64,061,260	390,601,010
3	Supporting team's fees	9,206,000	13,926,000	14,134,000	18,135,000	10,315,000	65,716,000
4	Transportation	5,060,000	3,410,000	3,630,000	2,640,000	1,870,000	16,610,000
5	Trainer Salaries	35,200,000	32,400,000	34,400,000	22,400,000	16,600,000	141,000,000
6	Miscellaneous	29,652,480	34,016,960	34,634,720	33,590,320	22,397,280	154,291,760
Total		184,211,085	193,335,140	188,001,360	185,378,260	125,638,540	876,564,385

Note. Source: Data from Training Center

Table 6. Cost and Benefit Summary of the Training Program

COST SUMMARY		
1	Training Delivery*	876,564,385.00
2	Participant's transportation cost	1,296,964,800.00
3	Participant salaries plus benefit $f(1/2 \times 5.096.106.100)$	2,548,053,050.00
Total Cost		4,721,582,235.00
*Cost of training deliveries is the expense of Training Center, an autonomous body outside DGT. However, to fairly present the ROI, this cost is included as program cost component.		
ANNUALIZED PROGRAM BENEFIT**		
1	Job performance	4,529,539,268.40
2	Time to complete the audit	4,262,754,153.60
3	Reduction in errors	4,523,428,033.20
Total Benefit		13,315,721,455.20
**Ideally, if new skills are acquired, there should be some value for the use of those skills in year two or even year three. However, for short term training programs, only the first-year values are used, thus requiring the investment to have an acceptable return in one-year time period. (Phillip, 1997)		

Figure 1 (a), (b), and (c). Frequency Distribution of Level 3 Evaluation: (a) hard skills, (b) soft skills, (c) total hard skills and soft skills.

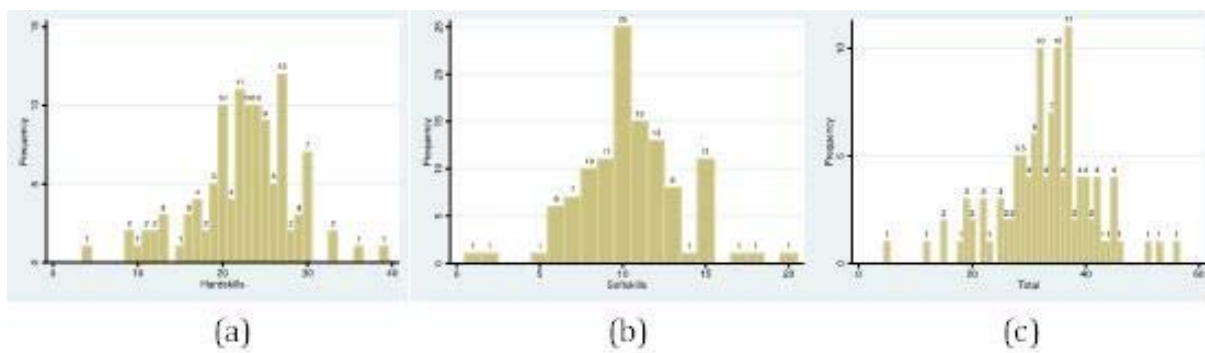


Figure 2. Frequency Distribution of Percentage Use of the Skills at Work

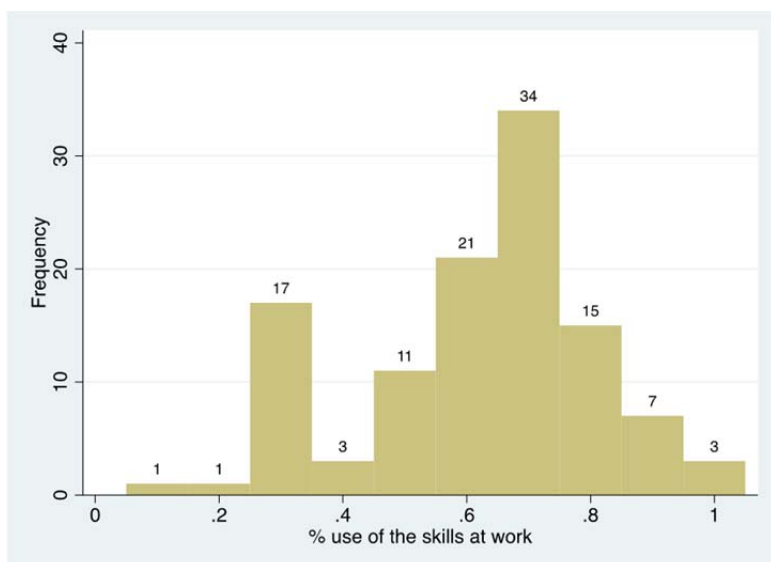


Figure 7. Pie Chart of the Increase in the Number of Audit Report after the Training

