

Measuring of Training Effectiveness (TREF)

Moges Mengstu Kassaw

Department of Research and Development, Ethiopian Ministry of Revenues

Tel: +2519-15-16-94-84 Email: mogmen25@gmail.com

Abstract

The objective of this study is to investigate and construct a valid and reliable instrument to measure training effectiveness or the development of Training Effectiveness (TREF). The research was done through three TREF measures including satisfaction, learning performance, and behavioral change measures. The development of TREF was applied using descriptive statistics average variance explained (AVE), Composite Reliability (CR), and reliability. Data were analyzed using reliability analysis, and structural equation modeling (SEM) using confirmatory factor analysis (CFA) through Stata and Jamovi software. Significantly, the TREF developed in this study can be utilized as a tool to evaluate the efficiency of private and public sector training. As a result, this effort will support the development of high-quality human resources in the private and public sectors for efficient policy implementation, as well as the enhancement of training programs and employee work performance.

Keywords: Training, TREF, CFA, Human resource development, training effectiveness

1. Introduction

The process of learning specific industry techniques is called training. It is the most important part of growing the business, and it helps the employees as well as the business grow and produce better products. An employee's skills, mindset, and behavior are systematically altered through training in order to perform a specific job (commercemates, 2022).

Training entails teaching a particular behavior or skill. When employees in a company need or want to learn new skills, they receive training (marketbusinessnews, 2023).

One of the most important ways to help people and businesses achieve their short- and long-term goals and objectives is through employee development and training. In addition to enhancing knowledge, skills, and attitudes, training and development offers a number of other advantages. According to (Nassazi, 2013), the following are typical advantages of employee training and development: 1) it boosts the motivation, self-assurance, and morale of employees. 2) Individuals are able to reduce waste, which reduces production costs. 3) It fosters a sense of safety, which in turn lowers employee turnover and absenteeism. 4) By providing employees with the skills they need to adapt to new and challenging circumstances, it increases their involvement in the change process. 5) It opens the door to promotion, recognition, and increased pay. 6) It contributes to the organization's improvement in staff availability and quality. It is important to remember that training and development programs improve people's skills and abilities, which leads to increased productivity (Bapna, R., et.al., 2013).

Human resources play a significant role among the resources required to achieve the institution's purpose. The institution should, to the greatest extent possible, provide the employee with task-oriented training, a pleasant working environment, and various benefits to enable the workforce to be effective. From this vantage point, it can be seen that organizations use a variety of research methods to determine the degree to which their employees are satisfied with their work and implement continuous improvement strategies to find and keep skilled, competitive, and professional employees.

Inland Revenue contributes significantly to the overall economic growth of a country; and so, the government is making a lot of efforts to continue the rapid development of the country. Therefore, the Ministry of Revenue of Ethiopia, to collect the national revenue required by the government and to prevent illegal trade, provides continuous training and various lessons to the employees.

In order to collect the national revenue required by the government and to prevent illegal trade, the Ministry of Revenue provides its employees with ongoing training and special education related to taxation and tax policy.

One employee is estimated to collect an average of two hundred fifty five thousand seven hundred birr (255,707) per day when we divide the institution's daily goal of collecting one billion one hundred and fifty million birr (1,150,684,931) by the total number of employees.

However, revenue collection will be seriously impacted if employees receive inadequate training to fill their knowledge and skill gaps. Since, it is essential to conduct this survey because it is necessary to know that the institution's training helps to collect income and builds employees' capacity.

This study attempts to investigate the effectiveness of the training provided by the Ministry of Revenues Tax Center as well as to identify the gaps seen in the training and to indicate the solutions to solve them. And the following research queries will be answered by the study: The training provided to the employees was effective? What should the Ministries and the training center do to make the training effective?

2. **Literature Review**

2.1. *Theoretical Review*

Human capital is considered a major asset in most developed countries and organizations; this is because it is believed that the skills, performance, and experience of employees are important assets in determining the present and future of the institution.

Different authors define the term training using different terms, (Mathis, RL, & Jackson, JH, 2011), training is the process by which people acquire skills to perform tasks and the focus of the training is work or activity. Such as (Jones, GR, and George, JM, 2005), description, training as an organized process in which people learn knowledge or skills for a specific purpose. Training is an ongoing and consistent effort by an organization or institution to expand education and knowledge among employees, enhance employee capabilities, and improve performance (Scott, G.,et.al., 2004).

The process of shaping and equipping employees through the addition of their skills, abilities, knowledge, and behavior is known as training (Ichsan, 2020). This enables work to be completed more quickly, effectively, and rationally. According to (Adnyani, N. L. P. R., & Dewi, A. A. S. K., 2019), employees will gain specific knowledge and be able to train skills that can later be used in the workplace if they receive training.

Training has become one of the crucial strategies to improve the performance of the workforce. Improved capabilities, knowledge, and skills of the talented workforce have proven to be a major source of capacity building to human resources for organizational productivity and competitive advantage in a global market, (Martin, 2015).

Helping employees feel satisfied and productive at work is one of the most important tasks any employer must perform. Employers are bound by the efficiency and quality of the employee's performance to achieve organizational goals; therefore, it is expected employers to meet various needs and create motivation for employees to feel satisfied with their work. (Tyson, S., and York, A , 2009).

Training is provided for four basic reasons, first new employees joining institutions or organizations are given training. The training provided to these employees is aimed at making employees familiar with the institution's mission, vision, rules and regulations, and the institution's operating systems. Secondly, existing employees will be trained to refresh and enhance their knowledge. Thirdly, as and when any improvements in technology are made, employees are trained to be aware of and adapt to these technological improvements. On the last, when promotion and career development are necessary, training is provided to prepare employees for higher job responsibilities and leadership.

2.1.1. *Training Effectiveness*

According to Albert J Lilly, (Lilly, 2011), monitoring employee training and measuring training effectiveness is a key objective of any training department. One of the ways to evaluate the effectiveness of employee training is by establishing Key Performance Indicators. When these institutional indicators are properly implemented and monitored, they serve as benchmarks for measuring and improving progress toward broader fundamental goals or objectives.

The effectiveness of training can be measured by the response level of trainees and/or the knowledge and learning that trainees gain from the training program; measuring the effectiveness of the training provided by institutions is one of the key elements in indicating how a job can be achieved appropriately. If the training program is proven to be effective, it will be more effective than the institution/organization needs. However, measuring training effectiveness is one of the biggest challenges facing institutions. Bramley (1994)

Overall training effectiveness is the extent to which the training objectives are achieved and benefited the Ministry and trainees, which can be evaluated using the combination of satisfaction (reaction), learning

performance, and behavioral change finding from the trainee’s feedback.

Effectiveness is evaluated on different four levels. All the results of the four levels are already merged and recoded into 'strongly agree', 'agree', 'neutral', “disagree” and “strongly disagree”. This section will summarize the result of all levels and this will provide a result of whether the provided training has been effective or not.

Measuring the effectiveness of training has many advantages (Ganesh M. & R. Indradevi, 2015), the first reason to conduct a training effectiveness survey is to measure how well the training went. Questions asked in the survey may reveal what participants liked or disliked about the training. It can show what improvements need to be made to make future training better or more relevant. Second, conducting training effectiveness surveys plays a vital role in making future trainings better and more effective. The omissions identified in the study and observed during the training will prevent trainers and organizers from making the same mistakes over and over again. A training effectiveness survey helps identify training gaps and areas and topics where employees need actual training. And a training effectiveness survey helps to determine whether the knowledge expected from the given training has been imparted to the trainees.

2.1.2. The Kirkpatrick evaluation model

Although there are various measurement methods and models to measure the effectiveness of training, many human resource development experts and researchers tend to By Dr. Donald Kirkpatrick (Dr. Donald Kirkpatrick), the model developed in 1959 (Kirkpatrick model) the main preferred model.

Kirkpatrick's model was a 4-step evaluation method developed in the late 1950s to measure the effectiveness of institutional training programs. It is preferred over other similar training measurement models to evaluate training effectiveness. This training effectiveness assessment framework is aimed at facilitating business globally and leveraging the knowledge gained from training in a cost-effective and time-efficient way to achieve better business results. Such as Kirkpatrick (Kirkpatrick model) According to the research theme, there are four steps to evaluate the effectiveness of training; these are:

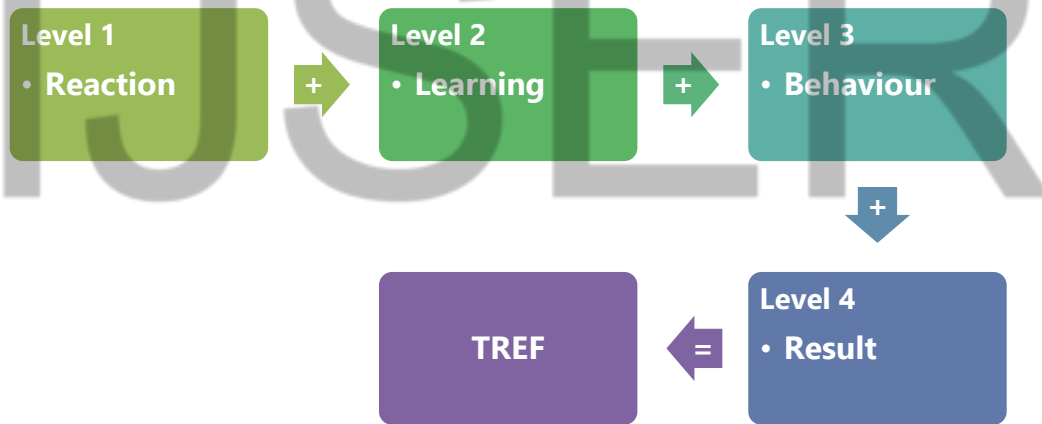


Figure 1 Kirkpatrick model

Source: Kirkpatrick Training Evaluation Model, 1979

Level 1. Satisfaction (Reaction):

The purpose of this evaluation point is straightforward, asking questions that confirm the trainees' opinions, allowing individuals to assess the relevance and usefulness of the training.

To measure and collect data, you can collect feedback from trainees about the training situation, fill out post-training questionnaires, and ask for reports from employees' immediate supervisors.

Level 2: Learning:

It helps to measure the skills and knowledge acquired as a result of the training. Learning is the assessment point we use to determine whether trainees have acquired the knowledge and skills intended by the training in line with the learning objectives set. The study methods varied from formal to informal tests and from self-assessment to group assessment. An assessment or evaluation is done to find out how much knowledge and understanding the

trainees had about the subject pre-training and post-training.

Level 3: Behavior:

To measure the extent to which the training has affected the performance or attitude of the trainee or participant, it helps to assess how well and accurately the participants apply the knowledge gained from the training on the job.

This level of measurement is evaluated by comparing the difference in the behavior of the trainee before and after the training after the completion of the training program. Behavior change is normally the immediate supervisor or thirdly side reviewer or by the individual as if it can be filled.

Level 4: Result:

It answers the question of what is the result or goal of the training. This phase focuses on evaluating the actual results of the training. This level, which is typically considered the main goal of the program, is to reduce training costs, achieve high returns on training costs, increase productivity, and reduce risk in the workplace, enabling more efficient operations and this scale is used to measure high work morale.

This is the effect in the workplace because of attending the training courses. Following the result, we asked whether the prepared training documents are helpful, was the training provided is directly related to your work, how relevant those contents of on workplace, and another important question was; has the training program helped you to do your job or any part of it better than before training?

2.2. *Empirical Review*

The impact of training on employee job performance was investigated by (Bin Atan, J., et.al., 2015), the research was carried out at a small and medium-sized business (SME) in Malaysia. The study looked at how the company's employees' job performance was affected by training. The study's functional area, which includes employee job performance and training, has been examined in relation to the company's efficient human resource management practices. The questionnaire-based survey was completed by 85 company employees from their respective production units. Effective training has a significant impact on employees' job performance, according to the study's findings.

A study conducted by (Al-Awawdeh, 2011), with the primary objective of statistically examining the connection between a training strategy and employee performance. The theoretical portion of the study identified key personnel's potential effect on performance as the study's goal. From Al-Bayt University's 651 administrative department staff, a sample of 120 employees was selected. The study demonstrated that Al-Bayt University employees' training needs must be determined using scientific methods, and training programs must be developed that focus on practice obstacles and work to overcome them. The study made a number of recommendations, one of which was to create training programs that take into account global developments and offer trainees training opportunities to prepare them for what might come from global changes, which may necessitate periodic retraining.

An integrated research model used by (Diamantidis, A., & Chatzoglou, P., 2014), combined the primary factors that have been shown to be related to training transfer with an examination of the relationship between training transfer and operational performance to examine the medium to long term effects of training programs on businesses. The study selected the training design, trainee self-efficacy, and work environment as transfer factors. Using data from 126 employees who have participated in various training programs in a number of Greek organizations, the validity of this model is evaluated using the structural equation modeling approach. The findings indicate that, in addition to trainees' self-efficacy and post-training behavior, the design of a training program has the greatest impact on job performance after training.

The primary objective of (Falola, H.,et.al., 2014), research was to investigate the impact of training and development on employee performance and the competitive advantage of an organization in the Nigerian banking sector. Two hundred and twenty-three valid questionnaires were filled out by selected banks in Lagos State, South-West Nigeria, using a simple random sampling technique for the descriptive research method. To meaningfully represent the raw data, descriptive statistics were used to analyze the collected data. The findings demonstrated a strong connection between employee performance, competitive advantage, and training and development.

2.3. *Conceptual Model*

This study is based on Kirkpatrick's research model to explore the effectiveness of training; this conceptual model is used to say whether the study is effective or not based on the three criteria of measurement (satisfaction or

reaction, learning performance, and attitude and behavioral change measures of the training).

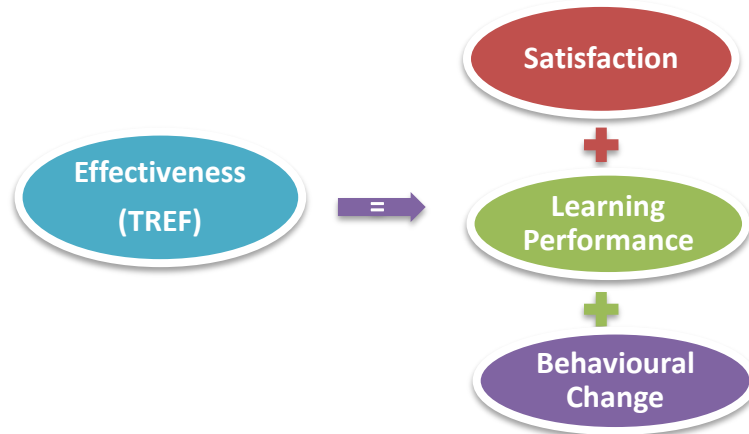


Figure 2 Training Effectiveness conceptual framework

Source: Author

3. Research Methodology

This study adopts a quantitative approach. The targeted group of study for this research was Ministry of Revenue employees attending training at the central level in 2020/2021;

The population of this study is 377 employees who have received training on tax laws and regulations from the branches of the Ministry of Revenue, and the sample was taken from the branch based on stratified sampling.

The sample size is a figure that is taken from the population or among the research elements to conduct a study and represents the majority. The overall sample size and sample of each stratum of the study were 194.

It will be difficult to obtain the required sample size using random sampling because of the workload in some departments, employees who are moving to educate taxpayers, and employees who remain outside the office for monitoring, control, and various tasks. In order to obtain the required sample size for the study, convenience sampling, also known as non-probability sampling, was used to collect the data.

Table 1 Sample size

Tax Center	Population	Sample
LTO	46	24
MTO	42	22
EAST	45	23
WEST	49	25
NORTH WEST	33	17
ADAMA	35	18
Head Quarter	127	65
Total	377	194

Source: MOR training center and Author Compilation

3.1. Data collection method

Convenience Sampling is a cost-effective, convenient, and easily accessible method of collecting data from representative data sources. This method of sample collection is quick, uncomplicated, and economical, making it the most common sampling method in many studies. The convenient sampling method was used to conduct this research, due to the nature of the work of the Ministry, it is difficult to find the selected employees in order to collect information at the desired time.

A written questionnaire to collect sufficient information regarding the primary source of information was prepared. The closed-ended section of the questionnaire is a quick and easy way for the informants and the researcher. Most of the close-ended questionnaires are prepared using a five-scored Likert scale. Accordingly, "1" indicates strongly disagree with the question or statement and "5" indicates strongly agree.

In this study, the questionnaire has two parts; Section one general information about the respondents focuses on questions like gender, age, education level, and work experience. Section two was developed to measure the variables of satisfaction (reaction), learning performance, and behavioral change. It comprised 20 closed-ended questions which utilized a five-point Likert scale. The items in the questionnaire were connected to respondents' perspective on training delivery, which consists of 3 question parts based on training measurement requirements; These are satisfaction (training preparation and delivery method), learning performance (improving knowledge and skills), and Behavioral change (shaping attitudes of the employees).

The questionnaire was prepared as closed-ended and open-ended, and the respondents were asked to fill in the provided answer space. The closed-ended section of the questionnaire is a quick and easy way for the informants and the researcher. Most of the close-ended questionnaires are prepared using a five-scored Likert scale. Accordingly, "1" indicates strongly disagree with the question or statement and "5" indicates strongly agree.

3.2. Data processing and analyzing

Since training effectiveness is a latent variable i.e., a variable we can't directly measure but gives rise to observe patterns in things we can observe, and then we chose the appropriate model Confirmatory Factor Analysis (CFA).

The TEF was developed and modified from the combination of (Kirkpatrick, 1959) and (AA Aziz, et.al., 2012), for evaluation of the training effectiveness provided by Ethiopian Ministry of Revenues Training Center. The TEF comprises three different sections to measure reaction (satisfaction), learning performance, and behavioral change; each dimension used in the TEF was operationally defined (see Table 2).

Table 2 Latent Variable Definition

Latent Variables	Item Measures	Description
TREF (Overall Training Effectiveness)	R1-B6	Overall training effectiveness is the extent to which the training objectives are achieved
Satisfaction (Reaction)	R1 – R10	Satisfaction (reaction) evaluation is the overall fulfillment towards the training design, material, delivery and quality, and personal satisfaction.
Learning Performance	L1 - L4	Learning performance evaluation is the improvement and acquired knowledge and skills in theoretical or practical.
Behavioral Change	B1 – B6	Behavior was measured by relationships with coworkers and customers.

Source: Author

In the first development of TREF, 10 positive statements were constructed for the satisfaction (reaction) dimension, 4 statements were constructed for learning performance and 6 statements were constructed for behavioral change measures, in sum, 20 items were developed.

Items for satisfaction were developed using training quality, design, content, and personal satisfaction, items for learning performance were developed using acquired knowledge and skill from the training, and behavioral change was measured as attitude change, being part of a team, and relationship with others.

3.3. Validity and Reliability of the Research

Data was collected at least three months after the completion of training. Content validity was applied to determine the validity of the questionnaires, and the distributed data collection questionnaires were reviewed by the training center department director and used as modifications to the final questionnaires applicable to the study.

In order to include all employees who takes training, the questionnaires were prepared in Amharic language, since it is official language of the country and translate into English for the study.

Prior to the actual data collection exercise, necessary adjustments were made on the basis of feedback given by certain employees and Ministries training center officials. After this, 190 questionnaires were collected from the expected sample (97.9%) out of the total questionnaires distributed.

There are various measures of reliability to determine the reliability of a study. One of the most commonly used

and common reliability measures is Cronbach's Alpha. Cronbach's Alpha is a measure used to assess the reliability or internal consistency of scales.

According to (Hair Jr, J. F., et.al., 2017), construct validity can be tested using convergent validity (the degree to which two measures of the same concepts are correlated), and nomological validity (all constructs are significantly correlated with each other). Hence, AVE and CR were assessed in the measurement model. Assessing validity and reliability, defined reliability as an assessment of the degree of consistency between multiple measurements of a variable.

This study assesses the consistency of the entire scale with Cronbach's alpha and its overall reliability of each factor of productivity values. All values yielded an alpha coefficient that exceeded the values of 0.70 suggested by (Hair Jr, J. F., et.al., 2017). From this result of Cronbach's alpha coefficient value, this questionnaire was accepted and admissible. To validate the instrument, this study also considered construct validation using analysis of structural model software (Stata and Jamovi) with maximum likelihood (ML) to analyze the data.

Jamovi and Stata version 15 statistical software package was used to analyze the collected data and significant findings. All the questionnaires were prepared in Amharic language and the data collected from the questionnaires were translated into English to be able to analyze the data by computer and the analysis was done.

Research ethics refers to the type of agreement that the person/group conducting the research enters into with the parties involved in the research. Since the study is based on the willingness of the employees, the purpose of the study is being explained to the participants, and if they do not want to give an opinion, it is not binding.

4. Results and Discussion

4.1. Respondent's profile

As shown in Table 3 below, 107, or 56.3%, of the trainees who participated in this study were male, while the remaining 83, or 43.7%, were female. Concerning the trainees' ages, 125, or 66.1% of the total workforce, are young workers between the ages of 18 and 35, followed by 48, or 25.4%, who are between the ages of 36 and 45. It employs 8.4% of workers over the age of 46. This suggests that the majority of ministry employees are young men. In terms of education, 98.4% of the 190 employees have a first degree or higher, 74.2% have a first degree, and 24.21 percent have a second degree or higher.

Table 3 respondent's profile

Variables	Freq.	Valid Percent
SEX		
female	83	43.68
male	107	56.32
AGE		
18-35	125	66.14
36-45	48	25.40
46-55	14	7.41
>55	2	1.06
EDUCATION		
TVET & Diploma	3	1.58
BA/BSc	141	74.21
MA/MSc & above	46	24.21

Source: Author

If the average result of the measuring scale (questionnaire) is below 3.39, it is a low result (Low), the average result is more than 3.4. If it is up to 3.79, it is considered moderate and if the average score is 3.8 and above, it is considered high (Pihie, Z. A. L., & Bagheri, A., 2013).

The respondents' average scores on the ten in-depth questions about the training and how it was prepared ranged from 2.73 to 4.03. The survey received the lowest score of 2.73. The four items (R7, R8, R9, and R10) that designate "the time given for the training," "training place," "accommodation," and "trainees overall satisfaction" have a low level of satisfaction (average score below 3.39).

Table 4 Measuring TREF using Weighted Average

TREF Measures	Variable symbol	Average	Interpretation
Satisfaction	R1	3.8388	High
	R2	3.7842	Medium
	R3	3.9973	High
	R4	3.9699	High
	R5	3.6694	Medium
	R6	4.0355	High
	R7	2.7295	Low
	R8	3.3443	Low
	R9	3.2322	Low
	R10	3.3798	Low
	Average	3.59809	Medium
Learning performance	L1	3.6421	Medium
	L2	3.7186	Medium
	L3	3.3251	Low
	L4	3.4071	Medium
	Average	3.523225	Medium
Attitude and Behavioral change	B1	3.7213	Medium
	B2	3.6831	Medium
	B3	3.7678	Medium
	B4	3.7705	Medium
	B5	3.7213	Medium
	B6	3.6311	Medium
	Average	3.71585	Medium

Source: Author

The average response score on the satisfaction level was 3.6, which is considered to be a medium level of satisfaction. The average result of the time allotted to the training (2.73) demonstrates that it is impossible to assert that not enough time is allotted to the training and that the time allotted to the pieces of training is sufficient. According to the trainees' responses to the training delivery questionnaires, which were prepared, the average score was 3.6 (medium level), indicating that the trainings had many strengths as well as many gaps.

The questionnaires that were administered to the trainees to determine whether or not they had acquired sufficient knowledge and skills from the training program revealed that, on average, they had been able to transfer the expected knowledge and skills. The participants in the study believe that the knowledge and skill gap they had as a result of receiving training has been reduced to some extent, regardless of whether the trainees have acquired the necessary knowledge and skills, according to the findings of the study. The respondents who stated that the training helped me become familiar with new procedures and technology received an average score of 3.32 and verbal interpretation at a low level in the detailed questions about the trainings' ability to impart knowledge and skills. The training that is provided typically has a medium level of knowledge and skill transfer (3.52). The survey found that the study's participants believe that the knowledge and skill gap they had as a result of receiving training has been reduced to some extent.

The training demonstrates that the trainees' average response scores range from 3.63 to 3.77, as shown in the table 4. The question, "The training has helped me feel satisfied with my work," received the highest mean score (3.77), while the question, "The training has developed a culture of teamwork," received the lowest mean score (3.33). The trainings played a moderate role in shaping employee attitudes and behaviors, as evidenced by their average score of 3.72 in influencing attitudes and behaviors. In general the training had a moderate impact on employees' attitudes and behaviors, as evidenced by the average score of 3.72 for attitudes and behavior shaping.

4.2. *Validity and Reliability*

The assessment of Average Variance Explain (i.e. AVE) for each construct was above .50 (Hair et al., 2017), as well as both reliability tests (i.e. Composite Reliability and Cronbach's Alpha) for each targeted construct being above .70 (Hair et al., 2017). Hence, the measurement model meets the validity criteria concept.

Table 5 validity and Reliability of Training Effectiveness

	AVE	CR	alpha	Satisfaction	Learning	Behavioral
Satisfaction	0.57	0.62	0.89	Satisfaction	1.0000	
Learning	0.66	0.85	0.82	Learning	0.7074	1.0000
Behavioral	0.73	0.91	0.83	Behavioral	0.6908	0.8018

Source: Author

The average AVE was 0.57 for satisfaction with ten items, 0.66 for learning performance with four items, and 0.73 for behavioral change measure with six items, as shown in table 5 above. As a result, there is convergent validation for each training effectiveness dimension with sufficient AVE and sufficient significance. Additionally, the construct reliability (CR) for behavioral change was 0.91, the CR for learning performance was 0.85, and the CR for satisfaction (reaction) was 0.62. This indicates that construct reliability is sufficient. The measurement model contained 20 items, all of which demonstrated an acceptable level of construct validity and construct reliability.

4.3. Confirmatory Factor Analysis (CFA)

The CFA model looks at how well three observed variables—satisfaction, learning performance, and behavior—measure the latent variable of training effectiveness. This is important because the level of training effectiveness is a difficult-to-measure complex latent concept.

We recorded four approximate model fit indices that each measure fits in a unique way to determine which of the three measurement models fits each data set best. First, the Tucker-Lewis index (TLI) and the comparative fit index (CFI), which measure the improvement in fit in comparison to a baseline model. Second, the standardized root mean square residual (SRMR) and the root mean square error of approximation (RMSEA), which measure the model's absolute fit to the data (Hoyle, 2000) and (Kline, 2011).

The best-fitting model was identified as the one with the lowest RMSEA and SRMR values but the highest CFI and TLI estimates. We highlighted all models with the simplest fit for data sets in which they fit equally well. To put it another way, all models were chosen when the fit statistics of the competing models were either equally high (i.e., CFI, TLI) or low (i.e., RMSEA, SRMR).

Model1: The first three-factor model with 20 items was represented (satisfaction was represented by items R1, R2, R3, R4, R5, R6, R7, R8, R9, and R10; Learning performance was represented by items L1, L2, L3 and L4; Behavioral change was represented by item B1, B2, B3, B4, B5, and B6) was analyzed by Stata Version15 revealed a poor fit with the χ^2 (df) = 477 (167), $p < 0.001$, SRMR = 0.058, CFI = 0.894, TLI = 0.879, RMSEA = 0.0988, indicating needs for further modification based on the Modification indices (M.I); a modification index (M.I) refers to the predicted decrease of the χ^2 statistic when a fixed parameter is freely estimated or an equality constraint is relaxed.

Table 6 Model 1 Fit Measures

Fit Measures						
χ^2	Df	P	CFI	TLI	SRMR	RMSEA
477	167	< .001	0.894	0.879	0.058	0.0988

Model 2: Five items (i.e. items R3, R6, R8, R9, and B4) were removed based on the Modification Indices of model 1. The three-factor model with 14 items (satisfaction was represented by items R1, R2, R4, R5, R7, R8, R9, and R10; Learning was represented by items L1, L2, and L4; Behavioral was represented by items B1, B2, B3, B5, and B6) was analyzed and showed a poor fit with the latent constructs (χ^2 (df) = 209 (74), $p < 0.001$, CFI = 0.929, TLI = 0.913, SRMR = 0.0488, and RMSEA = 0.0981), indicating further modification was necessary to improve the model fitness.

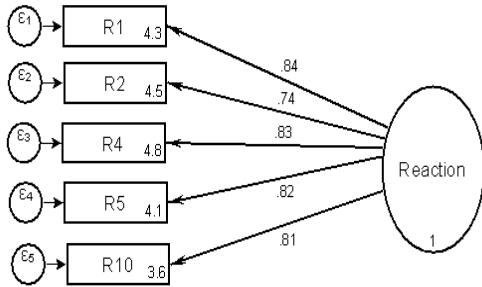
Table 7 Model 2 Fit Measures

Fit Measures						
χ^2	Df	P	CFI	TLI	SRMR	RMSEA
209	74	< .001	0.929	0.913	0.0488	0.0981

Source: Author

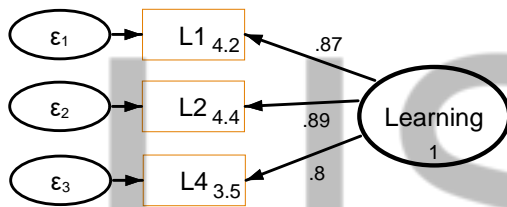
Model 3: Another two items (i.e. items R7 and L3) were removed based on the modification indices of model 2. The three-factor model with 13 items (satisfaction was represented by item R1, R2, R4, R5 and R10; Learning was represented by item L1, L2 and L4; Behavior was represented by item B1, B2, B3, B5, and B6) was analyzed and found a good fit with the latent (X2 (df) = 137.8 (62), $p < 0.001$, CFI = 0.962, TLI = 0.953, RMSEA = 0.0764 and SRMR = 0.032) and CD(coefficient of determination)=0.997. Standardized factor loadings showed that all the items in model 3 were well loaded on each latent construct.

Satisfaction:



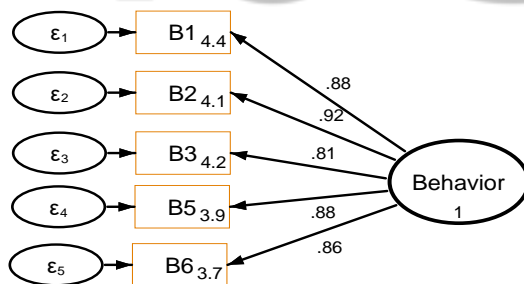
Fit statistic	Value
chi2_ms (5)	16.804
p > chi2	0.005
RMSEA	0.114
CFI	0.979
TLI	0.957
SRMR	0.026
CD	0.907

Learning:



Fit statistic	Value
p > chi2	0.000
RMSEA	0.000
CFI	.9999
TLI	.9999
SRMR	0.000
CD	0.897

Behavior:



Fit statistic	Value
chi2_ms (5)	33.567
p > chi2	0.000
RMSEA	0.175
CFI	0.966
TLI	0.932
SRMR	0.025
CD	0.945

Source: Stata-15 output and Author Compilation

Table 8 Model 3 Fit measures

Fit Measures										
									RMSEA 90% CI	
χ^2	df	p	CFI	TLI	SRMR	RMSEA	pclose	CD	Lower	Upper
137.83	62	0.000	0.962	0.953	0.032	0.0764	0.003	0.997	0.064	0.102

Source: Stata-15 output and Author Compilation

The research results show that the composite indicator from training effectiveness measures consisted of 3 main components and 13 indicators were fitted with the empirical data, determined from the Chi-square values = 137.83, comparative fit index (CFI) =0.962, Tucker-Lewis index (TLI) =0.953, standardized root mean square residual (SRMR) =0.032, root mean square error of approximation (RMSEA) =0.0764, and coefficient of determination (CD) =0.997.

4.4. Measurement Model for Training Effectiveness

The result of CFA analysis using correlated model in Figure 9 shows good results of GOF, i.e., $\chi^2/df=137.83$, root mean square error of approximation (RMSEA) = 0.083, standardized root mean square residual (SRMR) =0.032, comparative fit index (CFI) =0.962, Tucker-Lewis index (TLI)=0.953, and coefficient of determination (CD)=0.997.

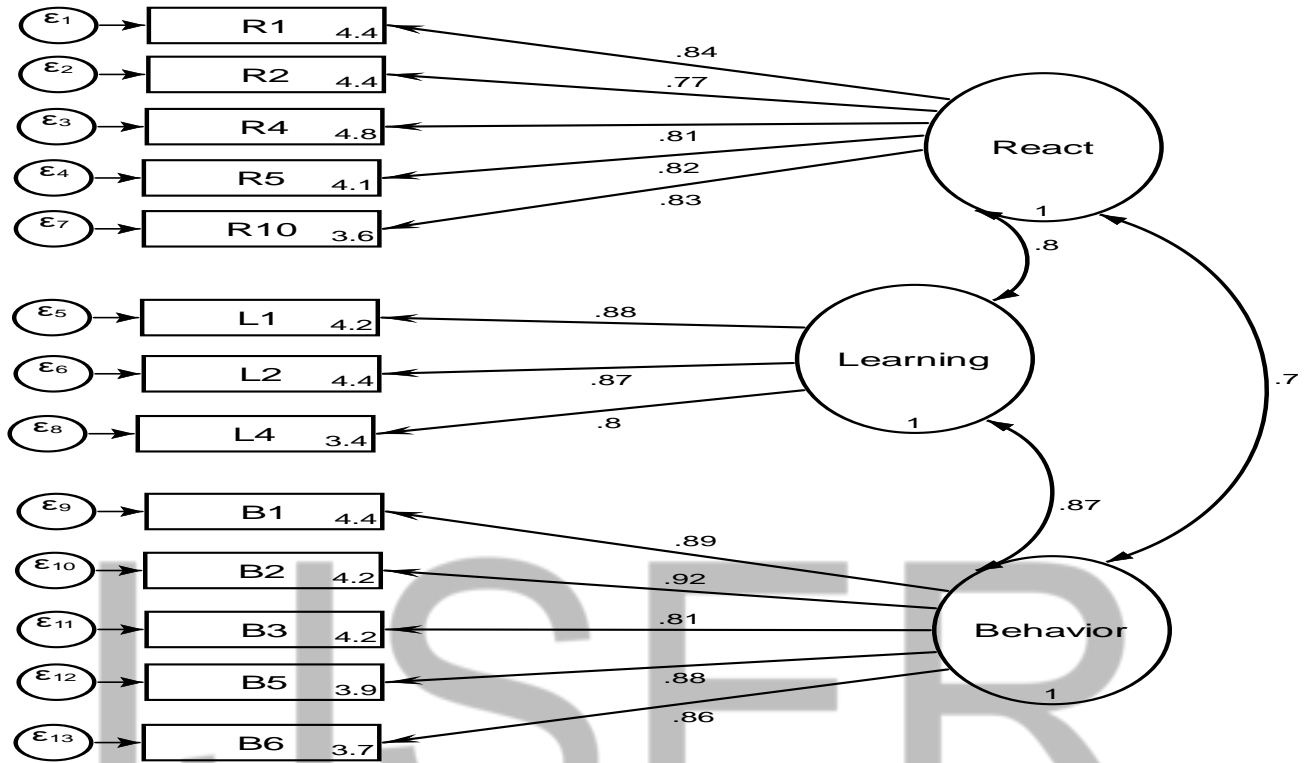


Figure 3 TREF Measurement Model

Source: Author

Table 9 Model 3 Fit measures

Fit Measures								
χ^2	df	p	CFI	TLI	SRMR	CD	RMSEA	pclose
137.829	62	0.031	0.962	0.953	0.032	0.997	0.083	0.003

Source: Stata-15 output and Author Compilation

The CFI, TLI, and SRMR do not find the models for complexity so it compares models in absolute terms. We expect that as model complexity increases (i.e., more indicators per factor), the performance of fit indices that include a penalty for model complexity will deteriorate. The RMSEA includes a drawback for model complexity, which will result in penalizing the multiple correlated factors model more than the higher-order model because the correlated factors model is less miserly (i.e., requires more parameters to be estimated) than the higher-order model.

4.5. Second-order Training Effectiveness Model

The 2nd Order CFA test is carried out by looking at the factor load value (>0.5). A factor load of 0.50 or more is considered vital enough to explain the latent construct (Hair, J. F., et.al., 2014).

The model shown in the Figure below is a figural representation of the structure of the training effectiveness as proposed by the author. For this “All Parameters Invariant” model, all aspects of the CFA model first and second-order factor loadings, factor variances, and unique and error variances were constrained to be equal across groups. The difference between the two models tests whether the factor structure imposed does or does not lead to degradation in the model fit.

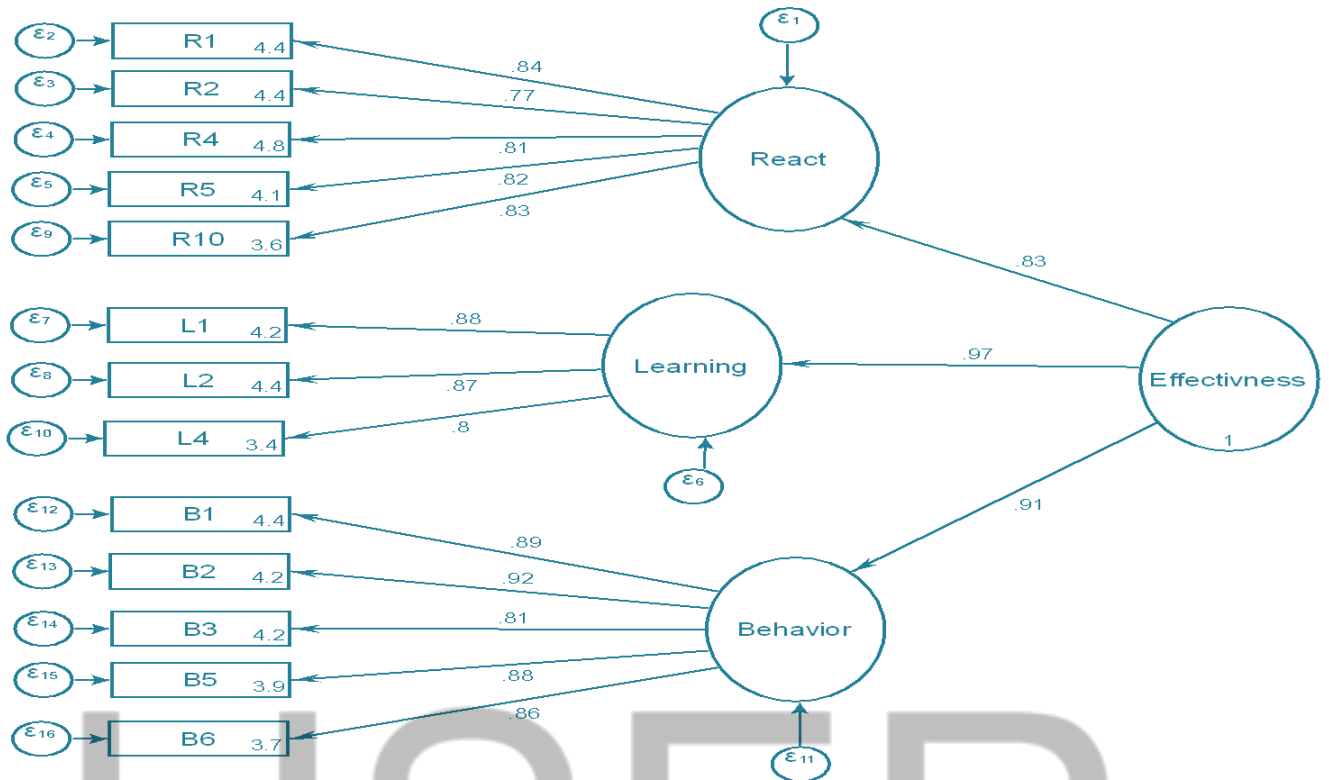


Figure 4 2nd order TREF Model
Source: Stata-15 output

Table 10 2nd order TREF Model Fit measures

Fit statistic	Value
chi2 ms (62)	137.829
p > chi2	0.062
RMSEA	0.063 (pclose 0.003)
CFI	0.971
TLI	0.964
SRMR	0.032
CD	0.974

Source: Stata-15 with Author Compilation

The structural model for TREF with acceptable Goodness of fit (GOF). The structural model shows that the most important construct of training effectiveness was learning performance (0.97), followed by behavioral change (0.91), and satisfaction or reaction measures (0.83). The finding is in line with the result of (Diamantidis, A., & Chatzoglou, P., 2014) and (Falola, H., et.al., 2014), the training has an impact on employee's job performance and learning performance.

5. Conclusions and Recommendations

The objective of this study was to evaluate the effectiveness of the training provided to employees of the Ministry of Revenue in 2020/2021, and construct a valid and reliable instrument to measure training effectiveness stretches by the Ministry of Revenues Ethiopia. The questionnaire was the main data collection method to get compulsory information for the study from employees and immediate supervisors. To determine the overall effectiveness of the training, Kirkpatrick's first three stages (reaction or satisfaction, learning performance, and behavioral change) were studied based on 20 detailed questions:

The research results show that most of the respondents were men aged between 18-35 years, and 98 percent and above of the total respondents have a bachelor's degree or higher. Out of the total employees who take training, 57% were male, and the remaining 43% are female employees. 78.8% of all trained workers are in the age group of 18-35.

To determine the overall effectiveness of the training, Kirkpatrick's first three stages (satisfaction or response, learning performance, and behavioral change) were studied based on 20 detailed questions, the measurement level (satisfaction or response) has a total average score of 3.6 (medium level), In terms of imparting knowledge and skills to employees (learning performance) it is a medium level (3.52), and the training in terms of shaping attitude and behavior (behavioral change) has the same medium result (3.72). In general, the result shows that the effectiveness of the training is at a medium level in the effort to evaluate the training effectiveness provided by the institution.

AVE and CR indicators from development consisted of 3 main components and 13 indicators were fitted with the empirical data, determined from the Chi-square values = 137.83, and no statistical significance ($p = 0.062$) at zero degrees of freedom ($df = 98$) indicated that model not different from the empirical data. In addition, found the RMSEA = 0.063, CFI = 0.97, TLI = 0.96 and SRMR = 0.032.

In general, it shows that all three levels (satisfaction, learning, and behavior) have moderate effectiveness according to the standards set to measure the effectiveness of training; the three-factor model with 13 items of the training effectiveness had a good fit and showed good maneuver values. It is a valid and reliable measurement to identify the effectiveness of training provided to tax officials across the Ministry of Revenue in Ethiopia.

Human Resource Development practitioners from public sectors are encouraged to use the TREF in training evaluation. TREF can be used to determine training effectiveness for any training program as a general evaluation. Since TREF included the evaluation of individual and organizational performance, TREF should be used at least three to five months after the completion of training. Specifically, this effort can support the improvement of quality human resources in the public sector for effective national policy implementation.

Additionally, TREF can also benefit the public service by offering a self-evaluation instrument to determine the effect of the training that they have provided. For researchers, TREF can be used in a survey to determine the relationship between training effectiveness and other variables, such as training place, accommodations, and employee norms to increase the effectiveness of training. It can also benefit the students that need to study training since there is a limited instrument to measure general training effectiveness.

In light of the study's findings, the following recommendation was made: Instead of being merely theoretical, the trainings that are provided ought to be hands-on, It is recommended that employees be able to attend the relevant topics for their work, and the training center should ensure that the right employee receives the training in terms of work behavior, It is recommended that the training be given by experts in the field who are close to the work and have practical knowledge and experience, to fill the trainees' knowledge and skill gaps, the training time should be increased.

Declaration:

Availability of data and materials: The data used during the study are available with the author and they are ready to make it available upon a reasonable request.

Competing interests: The authors declare there are no competing interests in this research paper.

Funding statement: The author received no financial support for the research, and/or publication of this article.

References

- AA Aziz, et.al. (2012). Developing a Structural Model of Assessing Students' Knowledge-Attitudes towards Sustainability,. *Social and Behavioral Sciences*, (<https://doi.org/10.1016/j.sbspro.2012.09.683>.), 515-519.
- Adnyani, N. L. P. R., & Dewi, A. A. S. K. (2019). Pengaruh Pengalaman Kerja, Prestasi Kerja Dan Pelatihan Terhadap Pengembangan Karier Karyawan. . *E-Jurnal Manajemen Universitas Udayana*, 8(7), 4073. <https://doi.org/10.24843/ejmunud.2019.v08.i07.p03>.
- Al-Awawdeh, W. (2011). Impact of Using Training Strategy on the Performance of Employees at the Administrative Departments at Al Al-Bayt University. *Al-Manarah for Research and Studies*, 78-106.
- Bapna, R.,et.al. (2013). Human capital investments and employee performance. *An analysis of IT services industry. Management Science*,, 641-658.
- Bin Atan, J., et.al. (2015). Impact of Training on Employees' Job Performance: A Case Study of Malaysian Small Medium Enterprise. *Review of Management*, 5(1/2), , 42-48.
- commercemates. (2022, January 15). *commercemates.com*. Retrieved from commercemates: <https://commercemates.com/what-is-training-types-methods-importance/>
- Diamantidis, A., & Chatzoglou, P. (2014). Employee Post-training Behavior and Performance: Evaluating the Results of the Training Process. *International Journal of Training & Development* , 159-166.
- Falola, H.,et.al. (2014). Effectiveness of Training and Development on Employees' Performance and Organization Competitiveness in the Nigerian Banking Industry. *Bulletin of the Transilvania University of Braşov*, 165-170.
- Ganesh M. & R. Indradevi. (2015). Importance and Effectiveness of Training and Development. *Mediterranean Journal of Social Sciences*. doi: 10.5901/mjss.2015.v6n1p334, 334-337.
- Hair Jr, J. F., et.al. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, 107-123.
- Hair, J. F.,et.al. (2014). *Multivariate data analysis*. Pearson new international edition. Essex: Pearson Education Limited.
- Hoyle, R. (2000). *In Handbook of Applied Multivariate Statistics and Mathematical Modeling*:. New York, NY, USA: Academic Press.
- Ichsan, R. N. (2020). Pengaruh Pelatihan terhadap Prestasi Kerja Karyawan pada PDAM Tirtanadi Cabang Padang Bulan Medan. *Jurnal Ilmiah METADATA*, 2(1), 72–75.
- Jayan, M., et.al. (2016). Investigating the relationship of adaptive leadership and leadership capabilities on leadership effectiveness in Sarawak schools. *Procedia-Social and Behavioral Sciences*, 540-545.
- Jones, GR, and George, JM. (2005). *Contemporary Management*. New York: McGraw Hills.
- Kirkpatrick, D. L. (1959). Techniques for evaluation training programs. . *Journal of the american society of training directors*, 21-26.
- Kline, R. (2011). *Principles and Practice of Structural Equation Modeling, 3rd ed*. New York, NY, USA: Guilford:

- Lilly, A. (2011). *Tracking Employee Training: Measuring training effectiveness with KPIs*. Canada: Robert Bacal/Bacal&Associates.
- marketbusinessnews. (2023, January 18). <https://marketbusinessnews.com/financial-glossary/training/>. Retrieved from marketbusinessnews.com: <https://marketbusinessnews.com/financial-glossary/training/>
- Martin, R. (2015). *How to begin with employee Training*. Retrieved from cleverism: <https://www.cleverism.com/ways-training-can-drive-employees-productivity/>
- Mathis, RL, & Jackson, JH. (2011). *Human Resource Management, 13th edition*. United States of America: South-western Cengage learning: West Publishing Company.
- Nassazi, N. (2013). *Effects of training on employee performance: Evidence from Uganda (Unpublished doctoral dissertation)*. Finland: University of Applied Sciences.
- Pihie, Z. A. L., & Bagheri, A. (2013). Self-efficacy and entrepreneurial intention: The mediation effect of self-regulation. *Vocations and Learning*, 385-401.
- Scott, G.,et.al. (2004). The effectiveness of creativity training: A quantitative review. *Creativity research journal*,16(4) , 361-386.
- Tyson, S., and York, A . (2009). *The Asia Foundation Training Needs Assessment Report. Essentials of HRM. 4th edition*. Vietnam: National Library of Vietnam.

IJSER

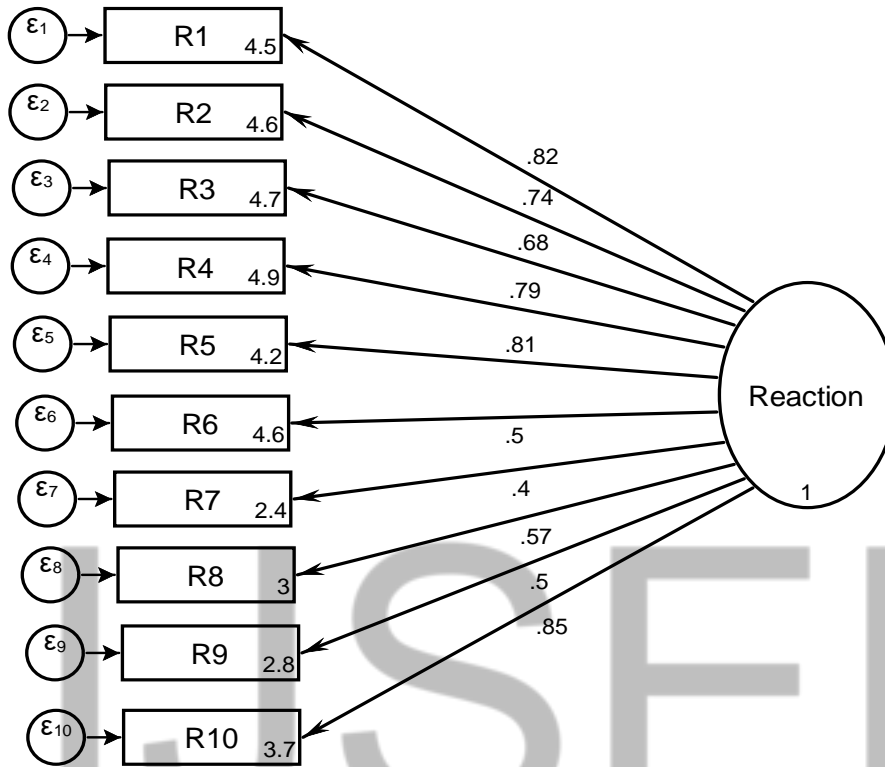
Appendix

1. Prepared questioners and coding

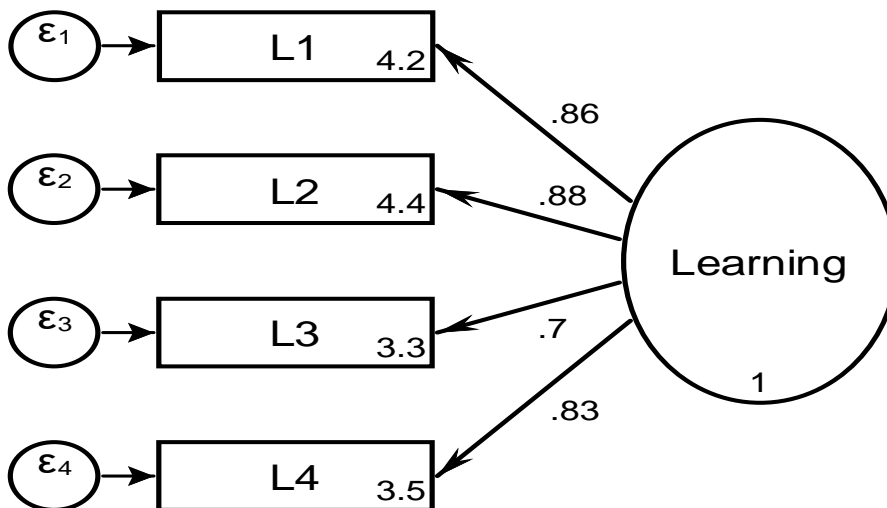
Reaction	CODE	Strongly disagree	disagree	neutral	agree	strongly agree
The training was supported by sufficient explanations and examples	R1					
The content of training documents is prepared in an understandable way	R2					
The prepared training documents are helpful for my work	R3					
The trainers have sufficient knowledge in the training they provide	R4					
The training method followed by the trainers was interesting.	R5					
The training given is directly related to my work	R6					
The time given for the training was sufficient	R7					
The place where the training was given was relatively comfortable	R8					
There was enough accommodation during training	R9					
In general, I found the training as expected	R10					
Learning						
The training allowed me to gain enough knowledge for my work	L1					
The training helped me to fill the knowledge and skill gap that I had	L2					
The training helped me to familiarize myself with new procedures/technology	L3					
The training helped me to be able to do my work independently without additional support	L4					
Behavior						
It helped me develop a sense of commitment to my work	B1					
The training has given me the confidence to do my job	B2					
The training helped me to treat customers politely	B3					
The training has developed the culture of teamwork	B4					
It helped me to carry out the responsibility given to me efficiently and effectively	B5					
The training helped me feel satisfied with my work	B6					

Model 1 Confirmatory factor analysis (CFA) for each dimension

Reaction



Learning



Behavior

