Measuring Training Effectiveness: Evidence from Malaysia

Siti Fardaniah Abdul Aziz

School of Psychology & Human Development Faculty of Social Sciences & Humanities
Universiti Kebangsaan Malaysia Email: daniah@ukm.my

Abstract

Global issues have emphasized the importance of training effectiveness as the essence of developing and managing quality human resources. Hence, to reinforce the management and implementation of national public policy for Malaysia, an easier way of measuring training effectiveness should be determined. Unfortunately, the information of psychometric properties for general instrument to measure training effectiveness was limited. Therefore, the objective of this article is to discuss research findings on constructing a valid and reliable instrument to measure training effectiveness or the development of General Training Effectiveness Scale (GTES). Research was done through three studies including jury validation, pilot study, and actual study*. The development of GTES's psychometric properties have included face validity, content validity, and constructs validity (including convergent, discriminant and nomological validity), as well as reliability and construct reliability. Data was analysed using reliability analysis, confirmatory factor analysis (CFA), and structural equation modelling (SEM) through SPSS and AMOS software. Significantly, the GTES as produced in this research can be used as an instrument to measure the effectiveness of training in private and public sector. Subsequently, this effort will support the improvement of quality human resources in public sector for effective national policies implementation.

Key words: Human resource development, training effectiveness, psychometric properties, General Training Effectiveness Scale

*The GTES development is a part of Phd research by Siti Fardaniah Abdul Aziz during her candidature at UPM with a title *"Mediation effect of training motivation on the relationships between trainee, training, and organizational characteristic support and training effectiveness"*. The GTES instrument is under protection of UKM's intellectual property rights.

1. Introduction

Training effectiveness is a crucial aspect to determine the returns on human resource development (HRD) investment (Noe, 2010; McGuire & Jorgensen, 2011; Werner and DeSimone, 2012). Interestingly, the four levels of training evaluation developed by Kirkpatrick (1959) tend to be the preferred framework to evaluate training effectiveness, either by researchers or practitioners

(Holton, 1996; Kirkpatrick and J. D. Kirkpatrick, 2010; Giangreco, Carugati, Denmark and Sebastiano, 2010). Despite the importance of measuring training effectiveness using the Kirkpatrick's (1959) model, general instrument to measure training effectiveness has received little attention.

Every year, a large amount of money was allocated for training investment. For example, Powell (2009) argues that U.S.A. spent as much as $134.39 billion annually on such programs; meanwhile, Griffin (2010) indicated that The National Employer Skills Survey in 2007 estimated that employers in the U.K. spent a total of £38.6 billion annually in training. In developing country, such as Malaysia, training is also a major concern, in which employers must annually contribute at least one per cent from their employees' salary for training fund (Chong, 2005; Abdullah and Mohd Yazam, 2009). Hence, training effectiveness can be seen as a global issue in developing and managing quality human resources.

According to Kirkpatrick (1959), training effectiveness can be evaluated using the four levels: reaction, learning, behavioral changes, and results. Ironically, findings by foreign studies, such as by Giangreco, Sebastiano and Peccei (2009) indicated that only large firms tend to use level three and four, however, most of small firms never gone beyond level one and two because of the complexity in evaluation process. This is also consistent with findings in national studies, such as by Chong (2005) that indicated from 106 manufacturing companies in Malaysia, 35% evaluated the reaction, 25% evaluated learning changes, 16.5% evaluated the behavioral changes, 11% evaluated results, and 12.5% did not organize any training evaluation. According to Kirkpatrick, the higher the level of training effectiveness the more difficulties, complexities, and expensive to measure it; this explains why it is very difficult to evaluate the complete levels of training effectiveness. Hence, there is a need to develop an instrument to measure general training effectiveness.

Therefore, the objective of this article is to discuss the development of general instrument to measure training effectiveness or General Training Effectiveness Scale (GTES). Without a general instrument to measure the effectiveness of training, it is difficult to determine the success of HRD program in the effort to produce quality human resources. Subsequently, it is difficult to reinforce the management and implementation of national public policy for Malaysia. Rationally, the GTES development can be used as a valid and reliable tool to determine the overall training effectiveness not only in public sector but also in private sector as well.

2. Literature Review

Overall training effectiveness (OTE) is the extent to which the training objectives are achieved and benefited for the company and trainees, which it can be evaluated using the combination of satisfaction, learning performance, individual performance, and organizational performance (Goldstein, 1986; Kirkpatrick, 1996; Holton, 2005; Tai, 2006; Bersin, 2008; Noe, 2010). There are various theories or models developed by previous research to explain on how to evaluate training effectiveness, in which the Kirkpatrick's (1959) model was seem to be the most preferred model used by researchers and practitioners in HRD (Griffin, 2010, Giangreco, et al. 2010). Meanwhile, some researchers including Cervero's (1988) and Quinones's (1997) demonstrated that each levels as proposed in the Kirkpatrick's (1959) model can be evaluated together to determine the overall training effectiveness.

Kirkpatrick's (1959) model of training evaluation is among the earliest models to evaluate training effectiveness and the preferred model for HRD researchers and practitioners, because it is simple and easy to understand (Noe, 2010; Griffin, 2010; Giangreco et al., 2010). However, many researchers, such as Alliger and Janak (1989), Cannon-Bowers, Salas, Tannenbaum and Mathieu (1995), Alliger, Tannenbaum, Bennett, Traver and Shetland (1997), Holton (1996), Bersin (2008), and Pineda (2010) criticized it. According to these researchers, its weaknesses include the following: (1) there is low correlation between levels; (2) each level of training evaluation does not relate to the achievement of the training's objectives; (3) the model is too simple and is more likely a taxonomy, which misses other operational measures; and (4) the levels should not be posited as a hierarchy. Consequently, many researchers have suggested alternative models to explain training effectiveness.

For example, Kraiger, Ford and Salas (1993) proposed the three domains of training outcomes to explain further about learning changes; this includes cognitive, skill-based, and affective outcomes. Meanwhile, Phillips (1997) popularized the technique of ROI and suggested it to be the fifth level of

training evaluation developed by Kirkpatrick (1959); and some researchers, such as Cannon-Bowers et al. (1995) and Holton (2005) proposed alternative model of training effectiveness that were more complex, which emphasized on the importance of training motivation. Ironically, although Kraiger et al. (1993) proposed a new perspective of learning changes, the model only explain further about individual/trainee changes or training evaluation level two and three in the Kirkpatrick's (1959) model. In fact, although Phillips (2003) suggested ROI as the fifth level in Kirkpatrick's model, however, it only explain further about the results or training evaluation level four in Kirkpatrick's model. This is consistent with Rowden (2005) and Bersin's (2008) arguments that ROI cannot verify training effectiveness if it is not proven that there was learning changes due to training.

Nonetheless, Cannon-Bowers et al. (1995) explained that training effectiveness was intervened by training motivation, and can be evaluated in four levels; this includes learning, training performance, job performance, and results. However, their training effectiveness model was developed specifically to test on the Navy trainees (more likely military training) as their research samples. Hence, the four levels of training effectiveness as proposed by them are not suitable for ordinary employee training program because in general training program, learning and training performance are actually the same dimension of training effectiveness. Meanwhile, Holton (2005) developed a more complex model to evaluate training effectiveness using human performance improvement as a basis for the model development, and he emphasized on the importance of training motivation. Holton redefined the next three levels of Kirkpatrick's (1959) model as learning performance, individual performance, and organizational performance.

Despite the fact that Holton (2005) and Cannon-Bowers et al. (1995) omitted the first level of training effectiveness in Kirkpatrick's (1959) model, however, Bersin (2008) argued that satisfaction or positive reaction towards training can be a powerful tool to predict training effectiveness. This is supported by a study from Ghosh, Joshi, Satyawadi, Mukherjee, and Ranjan (2011), in which, they did research on reaction evaluation solely as a measurement for training effectiveness and found that it can explain the effect of training on performance improvement and useful to determine the training program improvements. Meanwhile, Bersin (2008) argued that "satisfaction" is a more appropriate term to use than "reaction" because it distinguished the effect of training design and the positive reaction towards the training effectiveness. Therefore, as illustrated from the integration of training effectiveness models by Kirkpatrick (1959) and Holton (2005), the four measurements can be used to determine training effectiveness; these include satisfaction, learning performance, individual performance, and organizational performance.

Furthermore, Cervero's (1988) and Quinones's (1997) models of training effectiveness demonstrated that the four levels of training evaluation as proposed by Kirkpatrick (1959) can be considered together to determine the status of training effectiveness. This is consistent with Bersin (2008, pp. 62) arguments that one of the weaknesses of Kirkpatrick's (1959) model is that "the model incorrectly positions the levels as a hierarchy"; hence, the *levels* in most training effectiveness models should be referred to *dimensions*, and should be measured together to determine the status of training effectiveness. In fact, according to Noe (2010), the Kirkpatrick's (1959) model suggested that previous level would affect the next level of training effectiveness; however, the reaction and learning changes are measured in the same time, after the completion of training. Possibly, the popularity of Kirkpatrick's model explains why most researchers have tended to construct a hierarchy to explain the dimension of training effectiveness. For example, the respective cases of the third and fourth levels in the Cannon-Bowers et al. (1995) model and the second and third levels in Holton's (2005) model are suggested to be different levels, but can be measured at the same time sometime after training completion. Hence, the measurement of satisfaction, learning performance, individual performance, and organizational performance should be considered together as dimensions of training effectiveness.

In addition, the low relationships between each level of training effectiveness in Kirkpatrick's (1959) model, which was criticized by Holton (1996), have shown that the different levels in most training effectiveness models should be referred to as different dimensions. For example, Tai (2006) found a low correlation between *utility reaction* (satisfaction) and *learning* (learning performance) with $r = .21$ and $p = .05$; Tziner, Fisher, Senior and Weisberg (2007) found a moderate correlation between *instrumentality* (satisfaction) and *supervisor evaluation* (individual performance) with $r = .433$ and $p = .01$. Martineau (1995) found a low correlation between *reaction* (satisfaction) and *job performance after six months* (organizational performance) with $r = .26$ and $p = .01$. Axtell et al.

(1997) found a moderate correlation between *skills acquired* (learning performance) and *training transfer after one month* (individual performance) with $r = .49$ and $p = .01$; Chiaburu and Tekleab (2005) found a negligible correlation between *declarative knowledge* (learning performance) and *training maintenance* (organizational performance) with $r = .10$ and $p = .05$.

Furthermore, the effectiveness of training is not consistent if trainees show high training transfer or individual performance, but low learning or learning performance. For example: trainee A, from training program A, at organization A, had high learning performance and low individual performance; but trainee B, from training program B, at organization B, had low learning performance and high individual performance. In this case, based on the hierarchy, training program B is considered as more effective than training program A because the individual performance is higher; however, this is not consistent with the performance of previous training evaluation or learning performance. Other supporting aspects have been neglected; there is a probability that intensive follow up and intervention is provided in organization B, but not in organization A. In fact, many researchers, such as Kontoghiorghes (2004), Nijman, Nijhof, Wognum and Veldkamp (2006), and Nikandrou, Brinia and Bereri (2009) have stressed that organizational intervention, such as transfer-rewards can stimulate training transfer or individual performance. Hence, both individual performance and organizational performance can be measured together, but not as different level.

In addition, each satisfaction, learning performance, individual performance, and organizational performance measures different aspects and is used for different purposes. For example, Cannon-Bowers et al. (1995) measured *reaction*, or satisfaction, using satisfaction on content relevance, happiness, and overall reaction to training design; consistently, Ghosh et al. (2011) stressed that *reaction*, or satisfaction, is very useful for improving training programs in terms of planning, development, and organization. Meanwhile, Bersin (2008) argued that learning performance is used to measure the achievement of training objectives, which implies that learning performance is also used to determine the uses of training to improve the capability and qualification of an individual, especially to perform in his or her job.

Moreover, Chiaburu, Dam and Hutchins (2010) stressed that *transfer of training*, or individual performance, is used to measure whether what is learned from training is transferred to the workplace, which implies that individual performance is used to determine the uses of training to improve job performance. In addition, Bersin (2008) argued that organizational performance is used to measure the *total business and workgroup impact* due to training, which implies that organizational performance is used to determine the benefit or loss from the training program on organizational effectiveness. This demonstrated that most of researchers agreed that each level in the Kirkpatrick's (1959) model measures a different dimension and has its own purpose. Therefore, satisfaction, learning performance, individual performance, and organizational performance can be used as the dimension of training effectiveness.

Additionally, many HRD researchers and practitioners use only one selected level to evaluate training effectiveness and determine that training is effective. For example, to determine training effectiveness, Hand, Richards, and Slocum-Jr. (1973) focused only on *job performance* (individual performance), Baldwin, Magjuka, and Loher (1991) focused only on *learning performance* (learning performance), and Ghosh et al. (2011) focused only on *reaction to training* (satisfaction). Meanwhile, the Six Sigma approach focused only on *customer satisfaction* (Bersin, 2008) or a part of organizational performance. This shows that training is considered effective based on different dimensions. This strengthens the suggestion that satisfaction, learning performance, individual performance, and organizational performance can be measured together to determine the status of training effectiveness as a whole.

Furthermore, some researchers in human resource management (HRM) and HRD field of studies have evaluated various measurement of training effectiveness simultaneously to determine the degree of overall training effectiveness. These researchers have used multiple sources to collect data about various measurement of training effectiveness; however, they have combined those measurements to determine the contribution of HRM/HRD in organizational success. For example, Horgan and Muhlau (2006) did a study to determine the impact of HR system on employee performance among two different countries (Ireland and Netherland). The study used the combination of skills acquired (learning performance) and job performance (individual performance) as the measurement of training effectiveness in HR system. This highlights that these researchers have

tended to combine the various measurement of training effectiveness to determine the degree of training effectiveness as an evident to show the HRD's contribution on the organization's success.

In addition, Garcia (2005) investigated the relationship between training effectiveness and customer satisfaction, employee satisfaction, and shareholder satisfaction. The research have evaluated training effectiveness using various measurement including reaction (satisfaction), knowledge acquired (learning performance), knowledge applications (training transfer), and costs and benefits (organizational performance) simultaneously to determine the status of training effectiveness. This implies that precise training effectiveness could be measured by measuring those measurements simultaneously.

In summary, previous studies on training effectiveness have shown that training effectiveness seems to comprise four dimensions, which can be considered together to determine the status of training effectiveness. These include satisfaction, learning performance, individual performance, and organizational performance.

2.1. Dimensions of Training Effectiveness

Despite the fact that there were various terms used by previous researchers to measure different aspects of training effectiveness, these terms can be grouped into the four dimensions of training effectiveness; this includes satisfaction, learning performance, individual performance, and organizational performance (see Table 1). In addition, previous researchers have shown that all the dimensions were correlated with each other. This is consistent with arguments by Whetten (1989), Weick (1989) and Klimoski (1991) that all the variables involved should be correlated with each other to develop a theoretical framework.

Table-1. Examples of the different dimensions of overall training effectiveness

Training Effectiveness	Different terms used by previous researchers
Satisfaction	*Reaction* (Tannenbaum et al., 1991; Mathieu et al., 1992; Al-Ammar, 1994; Martineau, 1995; Colquitt et al., 2000; Tellis, 2004; Liao and Tai, 2006; Bell and Ford, 2007; Pilati and Borges-Andrade, 2008), *reaction to learning environment and content value* (Seyler et al., 1998), *reaction of post-training* (Tracey and Cardenas, 1996), *reaction of happiness, reaction of relevance, and overall reaction* (Cannon-Bowers et al., 1995), *course satisfaction* (Klein et al., 2006), *utility reaction* (Tai, 2006), *relatedness and instructional satisfaction* (Gegenfurtner, Festner et al., 2009), *training value* (Cheng and Ho, 1998), *transformed reaction* (Nease, 1999), *affective and utility reaction* (Tracey et al., 2001), *outcome expectancy* (Scaduto et al., 2008), and *instrumentality* (Tziner et al., 2007)
Learning performance	*Learning performance* (Mathieu et al., 1992; Martineau, 1995; Al-Ammar, 1994, Tai, 2006), *knowledge post-training* (Nease, 1999; Tracey and Cardenas, 1996), *declarative knowledge* (Colquitt et al., 2000; Tracey et al., 2001; Chiaburu and Tekleab, 2005), *skill acquisition* (Colquitt et al., 2000), *post-training self-efficacy* (Nease, 1999; Colquitt et al., 2000; Bell and Ford, 2007; Cunningham and Mahoney, 2004), *course grade* (Chuang et al., 2005; Klein et al., 2006; Liao and Tai, 2006), *test performance* (Tannenbaum et al., 1991), *skill acquired* (Axtell et al., 1997), *meta-cognition* (Klein et al., 2006), *academic test performance and physical performance* (Cannon-Bowers et al., 1995), *post-cognitive knowledge* (Myers, 1997), *learning outcomes* (Nijman et al., 2006), *training grades* (Tziner et al., 2007), and *application based knowledge* (Tracey et al., 2001)

Individual performance	*Training transfer* (Colquitt et al., 2000; Kontoghiorghes, 2004; Chiaburu and Tekleab, 2005; Switzer et al., 2005; Scaduto et al., 2008; Liebermann and Hoffmann, 2008), *training transfer after one month* (Axtell et al., 1997), *assessment performance* (Bell and Ford, 2007), *indirect training transfer such as professionalism, punctuality, consistency and team-building* (Nikandrou et al., 2009), *transfer outcomes* (Cheng and Ho, 1998; Nijman et al., 2006), *perceived training transfer* (Facteau et al., 1995; Hansen, 2001), *perceived training effectiveness* (Handy, 2008), *job performance after three months* (Martineau, 1995), *inspection and demerits* (Cannon-Bowers et al., 1995), *positive effect of training on work* (Pilati and Borges-Andrade, 2008), *perceived career development* (Bertolino et al., 2011), and *supervisor evaluation* (Tziner et al., 2007)
Organizational performance	*Training maintenance and training generalization* (Chiaburu and Tekleab, 2005; Scaduto et al., 2008), *job performance after six months* (Martineau, 1995), *job performance* (Colquitt et al., 2000), and *training transfer after one year* (Axtell et al., 1997)

As referred to Table 1, most of researchers used the general satisfaction, and satisfaction to training design, material, quality, and personal satisfaction as indicators to measure satisfaction; however, these indicators were interpreted differently. For example, general satisfaction was referred as *reaction of post-training* by Tracey and Cardenas (1996), *overall reaction* by Cannon-Bowers et al. (1995), *transformed reaction* by Nease (1999), and *reaction* by Tannenbaum, Mathieu, Salas and Cannon-Bowers (1991), Martineau (1995), Colquitt, LePine and Noe (2000), Liao and Tai (2006), Bell and Ford (2007), and Pilati and Borges-Andrade (2008). Meanwhile, satisfaction to training design was interpreted as *reaction to learning environment* by Seyler, Holton, Bates, Burnett, and Carvalho (1998), *instructional satisfaction* by Gegenfurtner, Festner, Gallenberger, Lehtinen and Gruber (2009), and *instrumentality* (Tziner et al., 2007). In addition, satisfaction to training material was interpreted as *reaction to content value* by Seyler et al. (1998), *utility reaction* by Tai (2006) and Tracey, Hinkin, Tannenbaum and Mathieu (2001), *relatedness* by Gegenfurtner, Festner et al. (2009), and *training value* by Cheng and Ho (1998). Additionally, satisfaction to training quality was interpreted as *course satisfaction* by Klein, Noe, and Wang (2006) and *outcome expectancy* by Scaduto, Lindsay and Chiaburu (2008). Finally, personal satisfaction was interpreted as *reaction of happiness* by Cannon-Bowers et al. (1995). This implies that satisfaction can be measured based on the satisfaction towards training material, design, quality, and personal satisfaction.

Further, learning performance is suggested to be measured using cognitive, skill, and affective outcomes by Kraiger, et al. (1993); hence, some researchers have been using *declarative knowledge, procedural knowledge,* and *meta-cognition* to measure learning performance. For example, declarative knowledge was referred as declarative knowledge by Colquitt et al. (2000), Tracey et al. (2001) and Chiaburu and Tekleab (2005), *learning performance* by Mathieu, Tannenbaum, and Salas, (1992), Martineau (1995), Al-Ammar (1994), and Tai (2006), *knowledge post-training* by Nease (1999), and Tracey and Cardenas (1996), *post-cognitive knowledge* by Myers (1997), *course grade* by Chuang, Liao and Tai (2005), Klein et al. (2006), and Liao and Tai (2006), *test performance* by Tannenbaum et al. (1991), *academic test performance* by Cannon-Bowers et al. (1995), *learning outcomes* by Nijman et al. (2006), and *training grades* by Tziner et al. (2007).

Procedural knowledge was referred as *skill acquisition* by Colquitt et al. (2000), *skill acquired* by Axtell et al. (1997), *application based knowledge* by Tracey et al. (2001), and *physical performance* by Cannon-Bowers et al. (1995). Meta-cognition was referred as *post-training self-efficacy* by Nease (1999), Colquitt et al. (2000), Bell and Ford (2007), and Cunningham and Mahoney (2004), and *meta-cognition* by Klein et al. (2006). Colquitt et al. (2000) found that declarative knowledge was correlated with procedural knowledge with $r_c = .18$ and $p = .05$; declarative knowledge was also correlated with post-training self-efficacy with $r_c = .19$ and $p = .05$; and procedural knowledge was correlated with post-training self-efficacy with $r_c = .24$ and $p = .05$. Tracey et al. (2001) found that declarative knowledge had a small effect with procedural knowledge with $\beta = .25$, $p = .01$. This implies that declarative knowledge, procedural knowledge, and meta-cognition are correlated with each other; hence, they are suitable indicators to measure learning performance.

Furthermore, most of researchers including Colquitt et al. (2000), Kontoghiorghes (2004), Chiaburu and Tekleab (2005), Scaduto et al. (2008), Liebermann and Hoffmann (2008), and Nikandrou et al. (2009) referred individual performance as *training transfer*. However, it is also referred using different terms, such as t*raining transfer after one month* by Axtell et al. (1997), *assessment performance* by Bell and Ford (2007), *transfer outcomes* by Cheng and Ho (1998) and Nijman et al. (2006), *perceived training transfer* by Facteau, Dobbins, Russell, Ladd and Kudisch (1995) and Hansen (2001), *perceived training effectiveness* by Handy (2008), *job performance after three months* by Martineau (1995), *inspection and demerits* by Cannon-Bowers et al. (1995), *positive effect of training on work* by Pilati and Borges-Andrade (2008), *perceived career development* by Bertolino, Truxillo and Fraccaroli (2011), and *supervisor evaluation* by Tziner et al. (2007). Most of researchers used job efficiency, effectiveness, and competencies as the indicators to measure individual performance; however, the correlation between these indicators was neglected and the information was unavailable from previous research.

Organizational performance was never really measured directly by previous researchers; however, it was measured indirectly using different terms. For example, it was referred as t*raining maintenance and training generalization* by Chiaburu and Tekleab (2005) and Scaduto et al. (2008), *job performance after six months* by Martineau (1995), *job performance* by Colquitt et al. (2000), and *training transfer after one year* by Axtell et al. (1997). As similar as individual performance, previous research have not provided any precise and adequate indicators to measure organizational performance. However, according to Kirkpatrick (1996), it can be measured using some indicators, such as teamwork improvement, customer satisfaction, and organization's goals achievement and reputation improvement. Hence, these indicators are suitable to measure organizational performance.

On the other hand, Table 2 shows previous research findings on the interaction between the four dimensions of training effectiveness including satisfaction, learning performance, individual performance, and organizational performance. As referred to Table 2, previous researches have demonstrated that each indicator for satisfaction had a positive correlation with each indicator for learning performance; this includes the relationship between content relevance, systematic training design, declarative knowledge, procedural knowledge, and meta-cognition. Each indicator for satisfaction also had a positive correlation with individual performance and organizational performance; this includes the relationship between content relevance, systematic training design, individual performance and organizational performance. Additionally, each indicator for learning performance, such as declarative knowledge, procedural knowledge, and meta-cognition also had a positive relationship with individual performance and organizational performance. Finally, it was also indicated that individual performance had a positive relationship with organizational performance.

Table-2. Examples of the interaction between different dimensions of overall training effectiveness

Interaction	Examples of previous research findings on the relationship between different dimensions of training effectiveness
Correlation between S and LP	• Tannenbaum et al. (1991) found $r = .32$, and Bell and Ford (2007) found $r = .46$ for the relationship between *general reaction* and *meta-cognition* • Tziner et al. (2007) found $r = .56$ for the relationship between satisfaction of *systematic training design* and *declarative knowledge* • Tracey et al. (2001) found $r = .44$ for the relationship between *satisfaction to systematic training design* and *procedural knowledge* • Klein et al. (2006) found $r = .31$ for the relationship between *satisfaction to systematic training design* and *meta-cognition*
Correlation between S and IP	• Scaduto et al. (2008) found $r = .58$ for the relationship between satisfaction to *systematic training design* and *individual performance* • Pilati and Borges-Andrade (2008) found $\beta = .25$, Martineau

		(1995) found $r = .45$ for the relationship between *general reaction* and *individual performance*
Correlation between S and OP	•	Scaduto et al. (2008) found $r = .53$, Martineau (1995) found $r = .26$ for the relationship between *general reaction* and *organizational performance*
Correlation between LP and IP	•	Tziner et al. (2007) found $r = .705$ for the relationship between *declarative knowledge* and *individual performance*
	•	Axtell et al. (1997) found $r = .49$ for the relationship between *procedural knowledge* and *individual performance*
	•	Bell and Ford (2007) found $r = .20$ for the relationship between *knowledge- acquired self-efficacy* and *individual performance*
Correlation between LP and OP	•	Chiaburu and Tekleab (2005) found $r = .10$ for the relationship between *declarative knowledge* and *organizational performance*
	•	Colquitt et al. (2000) found $\beta_s = .10$ for the relationship between *procedural knowledge* and *organizational performance*
Correlation between IP and OP	•	Chiaburu and Tekleab (2005) found $r = .76$, Scaduto et al. (2008) found $r = .51$, and Martineau (1995) found $r = .55$ for the relationship between *individual performance* and *organizational performance*. A significant effect between *training transfer* or IP and *job performance* or OP ($\beta_s = .59$) was found by Colquitt et al. (2000).

Notes: All figures are significant at least at 0.05 level of significant

S = satisfaction, LP = learning performance, IP = individual performance, OP = organizational performance

The synthesis on previous research shows the evidence that all the indicators involved to measure different dimensions of training effectiveness are correlated with each other. Table 2 also shows that the average correlation between each dimension of training effectiveness are only moderate; only some researchers, such as Chiaburu and Tekleab (2005), and Tziner et al. (2007) found a high correlation between different measurements of training effectiveness. This has shown that each dimension of training effectiveness can be measured together to determine the status of training effectiveness.

3. Methodology

To determine the psychometric properties for GTES, the current research used quantitative approach through three studies. Before organizing these studies, the GTES was developed using a self-construct by adapting questionnaire used by previous researches. The integration of Kirkpatrick (1959) and Quinones (1997) model of training effectiveness was adapted to construct the GTES. Then each dimension used in the GTES was operationally defined (see Table 3). The GTES comprises of four different sections to measure each satisfaction, learning performance, individual performance, and organizational performance. The GTES was developed and tested in Malay Version; however, for the purpose of reporting research outcomes, the English Version is used.

Table-3. Operational definition

Scale	Symbol	Description
Overall Training Effectiveness	OTE	Overall training effectiveness is the extent to which the training objectives are achieved and benefited for the company and trainees, which it can be evaluated using the combination of satisfaction, learning performance, individual performance, and organizational performance.
Satisfaction	S	Satisfaction evaluation is the overall fulfillment and happiness towards the training design, material, quality, and personal satisfaction.
Learning Performance	LP	Learning performance evaluation is the improvement or changes in declarative knowledge, procedural knowledge, and meta-cognition. Declarative knowledge includes knowledge about theories, facts, and method to use the training outcomes. Procedural knowledge includes the ability to use or apply training outcomes. Meta-cognition includes believes, certainty, and confidence in mastering knowledge and skills emphasized in training.
Individual Performance	IP	Individual performance evaluation is the improvement or changes in competencies, efficiencies, and effectiveness in workplace due to training outcomes. Competencies include the ability to solve work-related problems, produce better quality works, and fewer mistakes. Efficiencies include faster working; effectiveness includes the increase in total number of works.
Organizational Performance	OP	Organizational performance evaluation is the improvement or changes in teamwork, customer satisfaction, and organization's goals achievement and reputation improvement due to training outcomes.

In the first development of GTES, 10 positive statements were constructed for each dimension; in sum, 40 items were developed. Items for satisfaction were developed by adapting instrument used by Hansen (2001), Tellis (2004), and Klein et al. (2006). Items for satisfaction were developed using overall satisfaction and happiness towards training quality, design, content, and personal satisfaction. For learning performance, items were developed by adapting instrument used by Colquitt et al. (2000), Stanford (2000), and Hansen (2001). Items for learning performance were developed using three dimensions; this include declarative knowledge, procedural knowledge, and meta-cognition. Further, for individual performance, items were developed and adapted from Xiao (1996), Chiaburu and Tekleab (2005), and Handy (2008). Items for individual performance were developed using three dimensions; this include competencies, efficiencies, and effectiveness. Meanwhile, for organizational performance, items were developed and adapted from Chiaburu and Tekleab (2005) and Albrecht (2008). Items for organizational performance were developed using five dimensions; this include improvement in teamwork, customer/client satisfaction, achievement for organizational reputation, and achievement for organizational goals.

The first version of GTES comprises of 40 items; then, three studies were organized to refine the GTES. However, only 15 items left in the final study. From these 15 items, three items were selected to measure satisfaction, five items were selected to measure learning performance, four items were selected to measure individual performance, and three items were selected to measure organizational performance. Study 1 was organized through a jury validation as the early stage for the instrument development. Study 2 was organized through a pilot study involving 33 samples to prepare the instrument for actual research. Study 3 was organized through an actual research involving 281 respondents.

Study 1

Study 1 was a jury validation. In jury validation, three experts were appointed as juries for content validity process. These experts were lecturers in University Putra Malaysia (UPM) that came from various specialization of HRD, such as in research method, HRD policies, and training evaluation. These experts have consented to be appointed as juries. Using a quantitative approach, these experts play role as juries to validate the instrument by giving their scores from one to 10, as well as comments to improve on each items. Score one represent the opinion of "strongly disagreed" and 10 represent "strongly agreed" for these items. Using average scores, only items that have coefficient validity more than 0.7 were included in the GTES. This is consistent with accepted scores for reliability analysis. Some scholars, such as Hair, Black, Babin, Anderson and Tatham (2010), Pallant (2011), and DeVellis, (2012) suggested that the level of reliability that can be accepted is $\alpha \geq$ .70. Hence, the cut-off point was used to determine the content validity. Items that have less than 0.7 score is modified or eliminated from the scales. Based on study 1 only four items were selected for satisfaction, nine items were selected for learning performance, 10 items were selected for individual performance, and 7 items were selected for organizational performance. In study 1, data was analyzed using descriptive statistics, such as mean (average). The face validity and content validity was determined in study 1.

Study 2

Study 2 was a pilot test to prepare the GTES for actual research. The pilot study involved a number of 33 academic staffs in a faculty in a public university in Malaysia. These respondents used to attend various training programs organized by the Academic Development Centre (ADC) in the university. The GTES was distributed using formal e-mails to a number of 246 academic staffs at the faculty; only 33 were willing to get involved in this study and have returned the email. To get involved in the pilot test, these respondents have to remember about one training program that they used to attend. Then, they should answer the question honestly. They were given a week to return the email. In study 2, data was analysed using SPSS for reliability analysis.

Study 3

Study 3 was organized through actual research using a survey. The survey was organized to academic staffs in a public university in Malaysia. These academic staffs have attended one of 17 general training program organized by the ADC in the year 2011. ADC is a department in the university that provides training programs for academic staffs at the university. Usually ADC provides training programs that is useful for current and future job needs for these academic staffs whether offered as an option or mandated. The training programs included were:

(1) Basic Course of Research & Publication (Series 1)
(2) Metacognitive Teaching Method
(3) Basic Counselling Skill Course (Series 1)
(4) Research Method (Series 2)
(5) Effective Mentoring Strategy (Series 1)
(6) Basic Course of Internal Audit (Series 1)
(7) Higher Institutional Educational Teaching & Learning Course (Series 1)
(8) NVIVO Course (Series 1)
(9) Program Based Learning
(10) Effective Answering NCR Technique Workshop
(11) Basic Course of Research & Publication (Series 2)
(12) Rasch Model Winstep Course (Series 1)
(13) Integration of Web 2.0 in Teaching & Learning Course
(14) Effective Communication in Teaching & Learning
(15) NVIVO Course (Series 2)
(16) SPSS Course (Series 2)
(17) Learning Strategy

In most of the training, there were 40 participants in each training program. According to Buxbaum (1995), at least 50% sample is needed to evaluate training effectiveness for a training that

has lesser than 50 participants. Fortunately, 50% of participants for each training program were willing to participate in this study. In sum, only 284 academic staffs were willing to participate in this study from a population of 338 participants. A stratified simple random sampling method was organized to determine selected sample for each training program since some of the 284 participants attended more than one training programs. However, the sample size was reduced from 284 to 281 samples due to outliers. The 281 sample size have .99 value of power ($1-\beta$ error probability) when tested using G-Power software. According to Fritz and MacKinnon (2007) the value of power should be more than .80 to make sure an adequate sample size if using structural equation model (SEM).

In study 3, data was collected at least two to three months after the completion of training. Respondents were explained about the objective of this study at the beginning of training. If they are consent to get involved in the study, they have to fill questionnaire given to them by giving their formal university email address. Then, two to three months after the completion of each training program, the GTES questionnaire was emailed to them. To encourage participation, a lucky draw was organized, in which the ADC was responsible to organize the lucky draw. Winners were selected based on their email address. In this study, ADC had given their permission for data collection, as well as to help and assist data collection process. Meanwhile, to remain confidentiality and anonymity, respondents do not have to give their name or any identification that could reveal their personal details. They just have to give any email address that is active, in which, the email address is eligible for the lucky draw. They were also informed at the beginning of the training that participation in this research is based on volunteerism and confidential. Hence, by answering and returning the questionnaire, they are considered as agreed to participate in this research.

Data was analysed using SEM and independent sample t-test. An exploratory data analysis (EDA) including the normality test, linearity test, heteroscedasticity of errors, independent of errors, multicolinearity test, and outliers was organized; no violation of multivariate assumptions was found. Using SEM-AMOS, a confirmatory factor analysis (CFA) was organized to determine appropriate items for each dimension used in GTES. Then, the construct validity and construct reliability was determined.

4. Findings And Discussion

4.1. Face validity and Content validity

Face validity and content validity were determined during literature review and jury validation in study 1. Each dimension of GTES was operationally defined using literature review (see Table 3). Then, items were constructed using self-construct and adaptation from previous research. Forty items were constructed in the first version of GTES. Then, a jury validation was organized to refine the GTES to determine both face and content validity. During study 1, 30 items were selected for the second version of GTES. These items were refined based on comments from the juries; these juries have also given coefficient validity scores (see Table 4). The average coefficient value for satisfaction was 0.916, for learning performance was 0.896, for individual performance was 0.907, and for organizational performance was 0.890. The average coefficient value for each dimension of GTES was greater than 0.7 indicating that the GTES has an acceptable face and content validity level.

**Table-4.** Content validity for General Training Effectiveness Scale (GTES)

Items	Expert Panel 1	Expert Panel 2	Expert Panel 3	Average scores of content validity
Satisfaction				
OTE1	8	10	10	.933
OTE2	8	10	10	.933
OTE3	7	9	10	.867
OTE4	8	10	10	.933
Average of Coefficient Validity	.775	.975	1.0	.916

Learning Performance				
OTE5	8	10	10	.933
OTE6	6	10	10	.867
OTE7	6	10	10	.867
OTE8	6	10	10	.867
OTE9	7	10	10	.900
OTE10	7	10	10	.900
OTE11	8	10	10	.933
OTE12	7	10	10	.900
OTE13	8	9	10	.900
Average of Coefficient Validity	.7	.99	1.0	.896
Individual Performance				
OTE14	8	10	10	.933
OTE15	8	9	9	.867
OTE16	8	10	10	.933
OTE17	9	10	10	.967
OTE18	7	9	10	.867
OTE19	6	10	10	.867
OTE20	7	10	10	.900
OTE21	8	10	10	.933
OTE22	7	9	10	.867
OTE23	8	10	10	.933
Average of Coefficient Validity	.76	.97	.99	.907
Organizational Performance				
OTE24	7	10	9	.833
OTE25	6	10	10	.867
OTE26	6	10	10	.867
OTE27	8	10	10	.933
OTE28	6	10	10	.867
OTE29	9	9	10	.933
OTE30	9	9	10	.933
Average of Coefficient Validity	.72	.97	.98	.890

4.2. Confirmatory factor analysis

According to Hair et al. (2010, p. 773), CFA "is a way of testing how well measured variables represent a smaller number of constructs". Hence, CFA was organized for each dimension of GTES. Each dimension of GTES has acceptable goodness of fit (GOF) except for satisfaction because the number of items was not enough for AMOS analysis (see table 5). However, the GFI for satisfaction was 0.899 indicating that it almost has acceptable GOF. Hair et al. (2010) stressed that to indicate acceptable GOF, the GFI should be greater than 0.90 and the value for average variance extracted AVE should be greater than 0.5. Hence, satisfaction was included in the measurement model. In fact, Table 5 has shown the AVE for each dimension of GTES was sufficient. On the other hand, Figure 1 demonstrated the CFA for each dimension of GTES. These have shown that each dimension of GTES has acceptable CFA level; in which, there were four items for satisfaction, six items for learning performance, five items for individual performance, and four items for organizational performance.

Figure-1. Confirmatory factor analysis (CFA) for each dimension in GTES
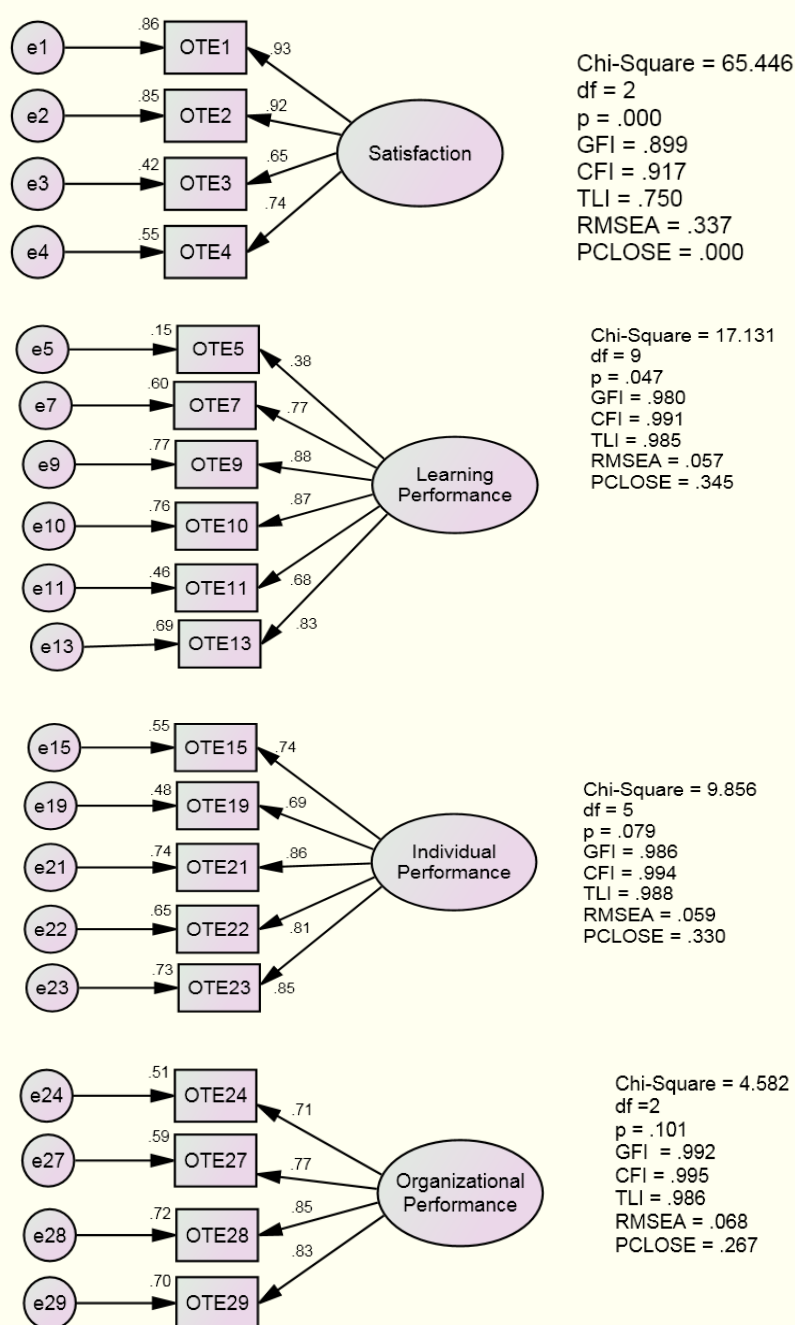
Notes: All regression weights and variances are significant at .0001 level of significant

Table-5. Goodness-of-fit indices for each dimension of GTES based on confirmatory factor analysis (CFA)

GOF index	CMIN (x^2)	df	GFI	CFI	TLI	RMSEA	AVE
Satisfaction	65.446 (p = .000)	2	.899	.917	.750	.337 (PCLOSE = .000)	0.67
Learning Performance	17.131 (p = .047)	9	.980	.991	.985	.057 (PCLOSE = .345)	0.57
Individual Performance	9.856 (p = .079)	5	.986	.994	.988	.059 (PCLOSE = .330)	0.63
Organizational Performance	4.582 (p = .101)	2	.992	.995	.986	.068 (PCLOSE = .267)	0.63

4.3. Construct Validity and Construct Reliability

"Construct validity provide confidence that items measures taken from a sample represent the actual true score that exists in the population" (Hair et al., 2010, pp. 777). Hair et al. (2010) suggested that construct validity can be tested using convergent validity (the degree to which two measures of the same concepts are correlated), discriminant validity (the degree to which two conceptually similar concepts are distinct), and nomological validity (all constructs are significantly correlated with each other). Hence, construct validity and construct reliability were assessed in the measurement model (see Figure 2); the formula was taken from Hair et al. (2010). The measurement model included items selected in CFA; the number of items for each dimension was reduced to get an acceptable GOF. The measurement model has acceptable GOF with $x^2(98) = 200.974$ with $p = .000$, $x^2/df = 2.050$, GFI = .912, CFI = .966, TLI = .959, and RMSEA = .061 with PCLOSE = .063.

Figure-2. Measurement Model for General Training Effectiveness Scale (GTES)
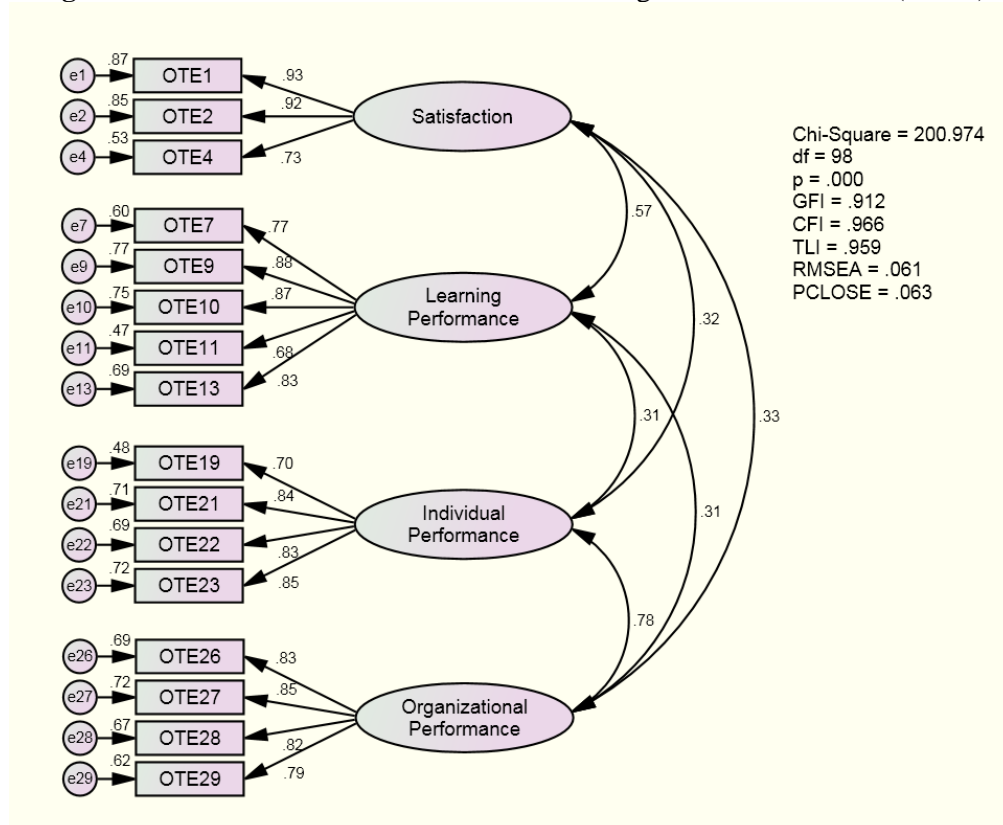
Notes: All regression weights, correlations, and variances are significant at .0001 level of significant

Figure 2 shows the measurement model, while Table 6 shows the value for AVE, correlation, squared correlation, and constructs reliability for each dimension in GTES. The AVE for satisfaction with three items was 0.75, for learning performance with five items was 0.656, for individual performance with four items was 0.65, and for organizational performance with four items was 0.675. Hence, each dimension in GTES has sufficient AVE with sufficient level of significant indicating a convergent validation. Meanwhile, the correlation between each dimension was also positively significant with $R > .3$ and $p = .0001$. This has indicated a nomological validation. In addition, the value for squared correlation between each constructs was lesser than the value of AVE for each construct indicating a discriminant validation. Additionally, the construct reliability for satisfaction was .898, for learning performance was 0.904, for individual performance was 0.882, and for organizational performance was 0.893. This has indicated sufficient level of construct reliability. In sum, there were 16 items in the measurement model, in which the measurement model had acceptable level of construct validity and construct reliability.

Table-6. Average variance extracted (AVE), constructs reliability (CR), correlation, and squared correlation among GTES constructs

AVE	CR		S	LP	IP	OP
.748	.898	**S**	-	.57	.32	.33
.655	.904	**LP**	.3249	-	.31	.31
.652	.882	**IP**	.1024	.0961	-	.78
.677	.893	**OP**	.1089	.0961	.6084	-

Notes: Values above the diagonal are correlations as produced by SEM. Values below the diagonal are squared correlation. All covariance are significant at .0001 level of significant

S = satisfaction, LP = learning performance, IP = individual performance, OP = organizational performance

4.4. Reliability

Reliability test was organized using Cronbach's Alpha in the pilot study (study 2) and actual study (study 3). The Cronbach's Alpha value for GTES in the pilot study was 0.833, meanwhile, for the actual study was 0.736 (see Table 7). The Cronbach's Alpha value for pilot study was tested based on 30 items on 33 samples; meanwhile the actual study was tested based on 15 items on 281 samples. The Cronbach's Alpha value for both studies was greater than 0.7 indicating that GTES has an acceptable reliability level. The reliability analysis has also shown that each item was significantly and positively correlated with each other.

Table-7. Reliability Analysis

Scale/ Variables	Cronbach's Alpha if item deleted	Cronbach's Alpha based on standardized items
GTES for pilot test (study 1) – Using 30 items		.833
Satisfaction	.877	
Learning Performance	.812	
Individual Performance	.797	
Organizational Performance	.826	
GTES for actual study (study 2) – Using 15 items		.736
Satisfaction	.661	
Learning Performance	.714	
Individual Performance	.636	
Organizational Performance	.620	

4.5. Final Version of General Training Effectiveness Scale (GTES)

The first version of GTES comprise of 40 items. However, it was reduced to 30 items for face and content validity. Then, the number of items was reduced to 19 items during CFA. During measurement model, the number of items has decreased to 16 items. Finally in the structural model, only 15 items were selected to be included in the final version of GTES (see Figure 3). The structural model for GTES has acceptable GOF with with $x^2(86) = 174.601$ with $p = .000$, $x^2/df = 2.030$, GFI = .928, CFI = .968, TLI = .961, and RMSEA = .061 with PCLOSE = .086. This has not much different with the measurement model (see Table 8). The AVE for satisfaction with three items was 0.746, for learning performance with five items was 0.656, for individual performance with four items was 0.65, and for organizational performance with three items was 0.67. Organizational performance was indicated to be the most important dimension, followed by individual performance, satisfaction, and learning performance.

Figure-3. Structural Model for General Training Effectiveness Scale (GTES)
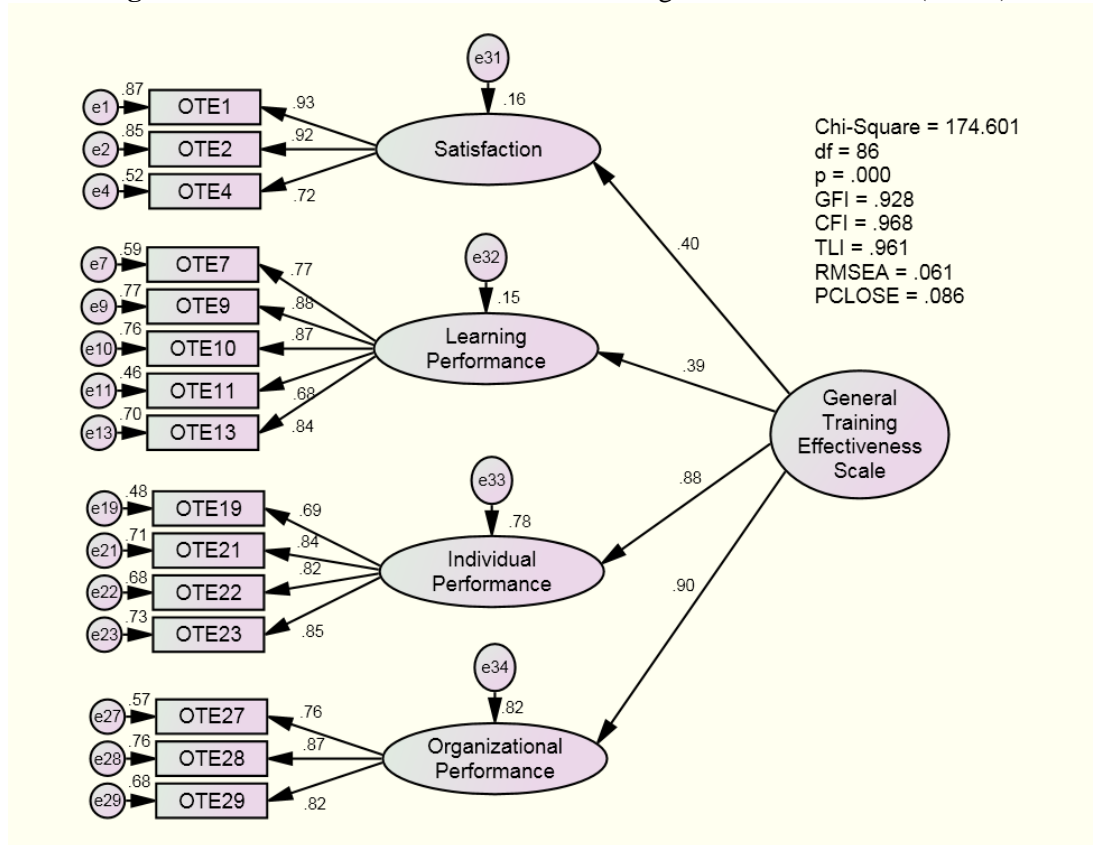
Notes: All regression weights and variances are significant at least at .05 level of significant

Table-8. Goodness-of-fit indices for the measurement and structural model

GOF index	CMIN (x^2)	df	GFI	CFI	TLI	RMSEA
Measurement Model	200.974 (p = .000)	98	.912	.966	.959	.061 (PCLOSE = .063)
Structural Model	174.601 (p = .000)	86	.928	.968	.961	.061 (PCLOSE = .086)

5. Conclusion And Recommendation

Training effectiveness is the essence of developing and managing quality human resources. Ironically, the information of psychometric properties for general instrument to measure training effectiveness was limited. Hence, the objective of this article was to discuss research findings on constructing a valid and reliable instrument to measure training effectiveness or the development of General Training Effectiveness Scale (GTES). Using the integration of Kirkpatrick (1959) and Quinones (1997) model of training effectiveness, GTES was constructed using four dimensions; these include satisfaction, learning performance, individual performance, and organizational performance. Three studies were organized to determine the psychometric properties for GTES including study 1 (jury validation), study 2 (pilot test), and study 3 (actual study). The first version of GTES comprised of 40 items; however, it was reduced to 15 items in the final version of GTES. The final version of GTES has passed some validity and reliability tests including face validity, content validity, CFA, construct validity, croanbach alpha reliability, and construct reliability. However, criterion-related validity cannot be assessed because of the nature of GTES. For example, GTES is an instrument to measure general training effectiveness that involves respondent from training programs. Hence, predictive validity cannot be done because it is impossible for the same group of respondent to attend the same training program in the future to test for predictive validity. Meanwhile, concurrent validity is difficult to organize because to date, a valid and reliable instrument to measure general training

effectiveness was never been reported yet. In fact, almost all training programs involved in this research (in study 3) had no academic test/ training performance test. Hence, data cannot be used to test for concurrent validity.

On the other hand, future researchers are recommended to test GTES using criterion-related validity especially for training programs that have a test/performance test at the end of training. Data from the training performance test can be used to test for concurrent validity. Future researchers are also encouraged to test the predictive validity in training programs that organized for the same respondent attended the same training programs. For example, GTES can be tested among school teachers that attended the same training every year. It is also recommended for future researchers to test GTES using other kind of sample since the current research only use sample from academic staffs working in a public university.

Further, HRD practitioner from public and private sectors are encouraged to use the GTES in training evaluation. GTES can be used to determine training effectiveness for any training programs as a general evaluation. With GTES, it is much easier to get overall picture whether training is effective or not. Since GTES included the evaluation of individual and organizational performance, GTES should be used at least two to three months after the completion of training. Specifically, this effort can support the improvement of quality human resources in public sector for effective national policies implementation.

References

Abdullah Lin & Mohd Yazam Sharif. 2009. Factors that affect training effectiveness at semiconductor wafer fabrication industry in Malaysia. *Journal of US-China Public Administration, 6*(4), 10-22.

Al-Ammar, S. A. 1994. *The influence of individual and organizational characteristics on training motivation and effectiveness.* Doctoral dissertation, State University of New York, USA (UMI No. 9504088).

Albrecht, V. L. 2008. *Determining the role of transfer implementation intent in predicting training transfer.* Doctoral dissertation, Capella University, USA (UMI No. 3304452).

Alliger, G. M., & Janak, E. A. 1989. Kirkpatrick's level of training criteria: Thirty years later. *Personnel Psychology, 42*, 331-342.

Alliger, G. M., Tannenbaum, S. L., Bennett, W. Jr., Traver, H., & Shetland, A. 1997. A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.

Axtell, C. M., Maitlis, S., & Yearta, S. K. 1997. Predicting immediate and longer-term transfer of training. *Personnel Review, 26*(3), 201-203.

Baldwin, T. T., Magjuka, R. J., & Loher, B. 1991. The perils of participation: Effects of the choice of training on trainee motivation and learning. *Personnel Psychology, 44*, 51-65.

Bell, B. S., & Ford, J. K. 2007. Reactions to skill assessment: The forgotten factor in explaining motivation to learn. *Human Resource Development Quarterly, 18*(1), 33-62.

Bersin, J. 2008. *The training measurement book*. San Francisco: Pfeiffer/John Wiley & Sons.

Bertolino, M., Truxillo, D. M., & Fraccaroli, F. 2011. Age as moderator of the relationship of proactive personality with training motivation, perceived career development from training, and training behavioral intentions. *Journal of Organizational Behavior, 32*, 248-263.

Buxbaum, A. 1995. *Training impact evaluation handbook*. Newton, MA: Family Planning Management Development, Management Sciences for Health.

Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. 1995. Toward theoretically based principles of training effectiveness: A model and initial empirical investigation. *Military Psychology, 7*(3), 141-164.

Cervero, R. M. 1988. *Effective continuing education for professionals*. San Francisco, USA: Jossey-Bass Publishers.

Cheng, E. W. L., & Ho, D. C. K. 1998. The effects of some attitudinal and organizational factors on transfer outcomes. *Journal of Managerial Psychology, 13*(5/6), 309-317.

Chiaburu, D. S., & Tekleab, A.G. 2005. Individual and contextual influences on multiple dimensions of training effectiveness. *Journal of European Industrial Training, 29*(8), 604-626.

Chiaburu, D. S., Dam, K. V., & Hutchins, H. M. 2010. Social support in the workplace and training transfer: a longitudinal analysis. *International Journal of Selection and Assessment, 18*, 187-200.

Chong, L. G. 2005. *Evaluating training effectiveness: An integrated perspective in Malaysia.* Unpublished doctoral dissertation, University of South Australia, Australia.

Chuang, A., Liao, W-C., & Tai, W-T. 2005. An investigation of individual and contextual factors influencing training variables. *Social Behavior and Personality, 33*(2), 159-174.

Colquitt, J. A., LePine, J. A., & Noe, R. A. 2000. Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*(5), 678-707.

Cunningham, G. B., & Mahoney, K. 2004. Self-efficacy of part-time employees in university athletics: The influence of organizational commitment, valence of training, and training motivation. *Journal of Sport Management 18*, 59-73.

DeVellis, R. F. 2012. *Scale Development: Theory and applications* (Vol. 26). California: Sage Publications.

Facteau, J. D., Dobbins, G. H., Russell, J. E., Ladd, R. T., & Kudisch, J. D. 1995. The influence of general perceptions of the training environment on pre-training motivation and perceived training transfer. *Journal of Management, 21*(1), 1-25.

Fritz, M. S., & MacKinnon, D. P. 2007. Required Sample Size to Detect the Mediated Effect. *Psychological Science, 18*(3), 233-239.

García, M. 2005. Training and business performance: The Spanish case. *International Journal of Human Resource Management, 16*(9), 1691-1710.

Gegenfurtner, A., Festner, D., Gallenberger, W., Lehtinen, E., & Gruber, H. 2009. Predicting autonomous and controlled motivation to transfer training. *International Journal of Training and Development, 13*(2), 124-138.

Ghosh, P., Joshi, J. P., Satyawadi, R., Mukherjee, U., & Ranjan, R. 2011. Evaluating effectiveness of a training programme with trainee reaction. *Industrial and Commercial Training, 43*(4), 247-255.

Giangreco, A., Carugati, A., Denmark, A., & Sebastiano, A. 2010. Are we doing the right thing? Food for thought on training evaluation and its context. *Personnel Review, 39*(2), 162-177.

Giangreco, A., Sebastiano, A. & Peccei, R. 2009. Trainees' reactions to training: an analysis of the factors affecting overall satisfaction with training. *The International Journal of Human Resources Management, 20*(1), 96-111.

Goldstein, L. 1986. *Training in organizations: Needs assessment, development and evaluation* (2 ed.). Monterey, CA: Brooks/Cole.

Griffin, R. P. 2010. Means and ends: Effective training evaluation. *Industrial and Commercial Training, 42*(4), 220-225.

Hair, J. F. J., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. 2010. *Multivariate data analysis* (7 ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Hand, H. H., Richards, M.D., & Slocum-Jr., J.W. 1973. Organizational climate and the effectiveness of a human relations training program. *Academy of Management Journal, June*.

Handy, L. A. W. 2008. *The importance of the work environment variables on the transfer of training.* Doctoral dissertation, North Carolina State University, USA (UMI No. 3306657).

Hansen, T. 2001. *A study examining the factors affecting training motivation.* Doctoral dissertation, Bowling Green State University, USA (UMI No. 3038434).

Holton, E. F., III. 2005. Holton's Evaluation Model: New evidence and construct elaborations. *Advances in Developing Human Resources, 7*(37), 37-54.

Holton, E. F., III. 1996. The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-21.

Horgan, J., & Muhlau, P. 2006. Human resource systems and employee performance in Ireland and the Netherlands: A test of the complementarity hypothesis. *International Journal of Human Resource Management, 17*(3), 414-439.

Kirkpatrick, D. 1959. Techniques for evaluating training programs. In Kirkpatrick, D. 1996. Great ideas revisited: Revisiting Kirkpatrick's four-level model. *Training and Development, 50*, 54-57.

Kirkpatrick, D. 1996. Great ideas revisited: revisiting Kirkpatrick's four-level model. *Training and Development, 50*, 54-57.

Kirkpatrick, D. L., & Kirkpatrick, J. D. 2010. *Evaluating training programs: The four levels* (3 ed.). USA: Berrett-Koehler Publishers.

Klein, H. J., Noe, R. A., & Wang, C. W. 2006. Motivation to learn and course outcomes: The impact of delivery mode, learning goal orientation, and perceived barriers and enablers. *Personnel Psychology, 59*, 665-702.

Klimoski, R. 1991. Theory presentation in Human Resource Management. *Human Resource Management Review, 1*(4), 253-271.

Kontoghiorghes, C. 2004. Reconceptualizing the learning transfer conceptual framework: empirical validation of a new systemic model. *International Journal of Training and Development, 8*(3), 210-221.

Kraiger, K., Ford, J. K., & Salas, E. 1993. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(311-328).

Liao, W.-C., & Tai, W-T. 2006. Organizational justice, motivation to learn, and training outcomes. *Social Behavior and Personality, 34*(5), 545-556.

Liebermann, S., & Hoffmann, S. 2008. The impact of practical relevance on training transfer: Evidence from a service quality training program for German bank clerks. *International Journal of Training and Development, 12*(2), 74-86.

Martineau, J. W. 1995. *A contextual examination of the effectiveness of a supervisory skills training program.* Doctoral dissertation, The Pennsylvania State University, USA (UMI No. 9600217).

McGuire, D., & Jorgensen, K. M. (Ed.). 2011. *Human Resource Development: Theory and Practice*. London: Sage Publication.

Myers, S. D. 1997. *The role of person, outcome, environmental, and learning variables in training effectiveness.* Doctoral dissertation, University of Tennessee, USA (UMI No. 9823114).

Nease, A. A. 1999. *Do motives matter? An examination of reasons for attending training and their influence on training effectiveness.* Doctoral dissertation, Rice University, USA (UMI No. 9969299).

Nijman, D. J. M., Nijhof, W. J., Wognum, A. A. M., & Veldkamp, B. P. 2006. Exploring differential effects of supervisor support on transfer of training. *Journal of European Industrial Training, 30*(7), 529-549.

Nikandrou, I., Brinia, V., & Bereri, E. 2009. Trainee perceptions of training transfer: An empirical analysis. *Journal of European Industrial Training, 33*(3), 255-270.

Noe, R. A. 2010. *Employee training and development* (5 ed.). Boston: McGraw Hill.

Pallant, J. 2011. *SPSS survival manual* (4 ed.). Australia: Allen & Unwin.

Phillips, J. J. 1997. *Return on investment in training and performance improvement programs* (1 ed.). USA: Butterworth-Heinemann.

Phillips, J. J. 2003. *Return on investment in training and performance improvement programs* (2 ed.). USA: Butterworth-Heinemann.

Pilati, R., & Borges-Andrade, J.E. 2008. Affective predictors of the effectiveness of training moderated by the cognitive complexity of expected competencies. *International Journal of Training and Development, 12*(4), 226-237.

Pineda, P. 2010. Perspective on practice evaluation of training in organisations: A proposal for an integrated model. *Journal of European Industrial Training 34*(7), 673-693.

Powell, J. L. 2009. *Transfer initiation and maintenance of training: Employees' perception of the relative influences of transfer intentions, general self-efficacy (GSE) and supervisor support.* Unpublished doctoral dissertation, Chesterfield University, Falls Church, Virginia, USA.

Quinones, M. A. 1997. 'Contextual influences in training', in M. A. Quiñones and A. Ehrenstein (Eds.), *Training for a Rapidly Changing Workforce* (pp. 177-99). Washington, DC: American Psychological Association.

Rowden, R. W. 2005. Exploring methods to evaluate the return-on-investment from training. *Business Forum, 27*(1).

Scaduto, A., Lindsay, D., & Chiaburu, D.S. 2008. Leader influences on training effectiveness: Motivation and outcome expectation processes. *International Journal of Training and Development, 12*(3), 158-170.

Seyler, D. L., Holton-III, E. F., Bates, R. A, Burnett, M. F, & Carvalho, M. A. 1998. Factors affecting motivation to transfer training. *International Journal of Training and Development, 2*(1), 2-16.

Stanford, M. A. 2000. *Individual and situational influences on pre-training motivation: Its impact on training motivation.* Doctoral dissertation, North Carolina State University, USA (UMI No. 9994056).

Tai, W.-T. 2006. Effects of training framing, general self-efficacy and training motivation on trainees' training effectiveness. *Personnel Review, 35*(1), 51-65.

Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. 1991. Meeting trainees' expectations: The influence of training fulfilment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology, 76*(6), 759-769.

Tellis, J. Y. T. S. 2004. *Relationships of individual, situational, motivational, training reaction factors, and motivation to transfer training.* Doctoral dissertation, Auburn University, USA (UMI No. 3136017).

Tracey, J. B., & Cardenas, C. G. 1996. Training effectiveness: An empirical examination of factors outside the training context. *Journal Research Journal, 20*(2), 113-123.

Tracey, J. B., Hinkin, T. R., Tannenbaum, S. I., & Mathieu, J. E. 2001. The influence of the individual characteristics and the work environment on varying levels of training outcomes. *Human Resource Development Quarterly 12*(1), 210-214.

Tziner, A., Fisher, M., Senior, T., & Weisberg, J. 2007. Effects of trainee characteristics on training effectiveness. *International Journal of Selection and Assessment, 15*(2), 167-174.

Weick, K. E. 1989. Theory construction as disciplined imagination. *Academy of Management Review, 14*(4), 516-531.

Werner, J. M., & DeSimone, R. L. 2012. *Human Resource Development 6e* (6 ed.). USA: Thomson South-Western.

Whetten, D. A. 1989. What constitutes a theoretical contribution? *Academy of Management Review, 14*, 490-495.

Xiao, J. 1996. The Relationship Between Organizational Factors and the Transfer of Training in the Electronics Industry in Shenzhen, China. *Human Resource Development Quarterly, 7*(1), 55-73.