

Evaluating the Impact of Individual Training on Unit's Operational Performance

Dr. Jay Brimstin
U.S. Army, Maneuver Center of Excellence
Ft Benning, GA
Jay.a.brimstin.civ@mail.mil

Toumnakone Annie Hester
U.S. Army, Maneuver Center of Excellence
Ft Benning, GA
Toumnakone.hester.ctr@mail.mil

ABSTRACT

The evaluation of training programs has long been considered one of the critical components of the analysis, design, develop, implementation, and evaluation (ADDIE) process. Typically though, training evaluations at best measure learning, with no measure of the learner's ability to apply what they learned on the job and no measure of the impact on organizational performance. This is especially true for military training programs designed to prepare soldiers for combat operations, as it is difficult to trace the results of the training program to organizational performance. This paper describes a project that implemented a robust training evaluation of an Army institutional training program that was designed to better prepare soldiers in units to deal with the improvised explosive devices (IED) threat. The Dismounted Counter IED Tactics Master Trainer (DCT-MT) course provided a unique opportunity to pilot an evaluation method using Kirkpatrick's four-level model for an institutional training course. The course's mobile training team conducted training for soldiers from the same brigade combat team prior to their deployments to Afghanistan. The evaluation employed a mixed method design to evaluate Levels I and II during the training and Levels III and IV during and after the unit's deployment. The findings from this evaluation that spanned a two-year period showed positive soldier reaction to the learning experience and significant improvement in their knowledge. Soldiers applied their knowledge in their units under certain conditions. A comparison of specific variables between units that did and did not receive the DCT-MT training showed a positive impact on the units' operational performance. This evaluation project demonstrated the ability to quantify the impact that training had on units' operational performance in a combat environment by using Kirkpatrick's model, and provides an evidence based approach for conducting future training program evaluations.

ABOUT THE AUTHORS

Dr. Jay Brimstin is the Deputy Director for the Directorate of Training and Doctrine (DOTD) at the Maneuver Center of Excellence (MCoE) at Fort Benning, Georgia. The Directorate is responsible for developing doctrine and training products, strategies, and programs for use throughout the MCoE and brigade combat teams across the Army. Dr. Brimstin has a PhD in education, specializing in training and performance improvement, as well as an MBA. Dr. Brimstin is a retired Army veteran with 30 years of service in which he served in multiple leadership and instructor roles. Included in his role as the Deputy Director for the DOTD is to function as the agency career program manager for career program 32, which is the Army's career program for Department of the Army Civilian training, capability, and doctrine warfighting developers. Dr. Brimstin's specialization in training and performance improvement, and his academic background in business provide him with a unique perspective on assessing the overall effectiveness of training programs on the Army's bottom line, fighting and winning the nation's wars.

Toumnakone Annie Hester is the lead evaluator for the Design and Evaluation team for the Dismounted C-IED Tactics Master Trainer course. She developed the evaluation plan for the course and has spent the last two years collecting and analyzing data, conducting interviews and focus group, and authored the final report. Her education includes a Masters of Public Administration and a Masters of Education in Training and Development. She's currently working on her PhD in Management and Organizational Leadership and is in the process of completing her dissertation. She has extensive experience in training deploying units as a member of the NC Army National Guard Pre-mobilization Training Assistant Team and future leaders of the Army as an ROTC instructor at East Carolina University. Her last deployment as a company commander in a combat Engineer Battalion, whose sole job was to look for IEDs, provided her with firsthand knowledge of the devastating effects of IEDs and the dynamic IED environment. Annie's experience in training, academic background, and deployments provided her with distinctive insights, which were added value in the evaluation process and in reporting the evaluation findings.

Evaluating the Impact of Individual Training on Unit's Operational Performance

Dr. Jay Brimstin
U.S. Army, Maneuver Center of Excellence
Ft Benning, GA
Jay.a.brimstin.civ@mail.mil

Toumnakone Hester
U.S. Army, Maneuver Center of Excellence
Ft Benning, GA
Toumnakone.hester.ctr@mail.mil

The evaluation of training programs has long been considered one of the critical components of the instructional design process. Typically though, training evaluations at best measure learning, with no measure of the learner's ability to apply what they learned on the job and no measure of the impact on organizational performance (Patel, 2010). This is especially true for military training programs designed to prepare soldiers for combat operations, as it is difficult to trace the results of the training program to organizational performance. This paper describes a project that implemented a robust training evaluation of an Army institutional training program that was designed to prepare soldiers in units to deal with the improvised explosive device (IED) threat. The Dismounted Counter IED Tactics Master Trainer (DCT-MT) course provided a unique opportunity to pilot an evaluation method using Kirkpatrick's four-level model for an institutional training course. The findings from this evaluation that spanned a two-year period showed positive soldier reaction to the learning experience and significant improvement in their knowledge. Soldiers applied their knowledge in their units under certain conditions. A comparison of specific variables between units that did and did not receive the DCT-MT training showed a positive impact on the units' operational performance. This evaluation project demonstrated the ability to quantify the impact that individual training had on units' operational performance in a combat environment by using Kirkpatrick's model, and provided an evidence based approach for conducting future training program evaluations.

BACKGROUND

The purpose of the DCT-MT course is to prepare soldiers to integrate training, advise leaders on threat assessments, and plan tactical employment of handheld devices to counter-IED related threats targeting dismounted patrols. The course was developed based on observations that units were lacking in their understanding of and ability to employ the multiple devices that were developed for detecting IEDs and they were not proficient in the small unit tactical techniques that had been identified as most effective. The course was implemented in 2012. The instructors trained soldiers to effectively assist leaders with training, planning, executing, and supervising at company and battalion levels. The ten day DCT-MT course was structured with the first week as classroom instruction and practical exercises. The second week was field-based situational training exercises (STX) and focused on the tactical employment of the dismounted C-IED handheld devices and a training assessment of the soldiers.

A fundamental question that arose during the development of the course was whether the training would make a difference in terms of unit performance and reducing casualties to dismounted units resulting from IEDs. This question of impact on organizational performance is a question that organizations have been struggling with for years as they invest billions of dollars annually into training programs (Bober & Bartlett, 2004). While many organizations conduct some level of evaluation for their training programs, most do not evaluate the degree to which individuals are able to apply what they learned on the job, and more importantly, whether the training resulted in improved organizational performance.

One of the many reasons why most organizations do not evaluate beyond Levels I (reaction) and II (learning) is because with each progressive level, the process becomes more time consuming and requires more resources. Moreover, demonstrating the cause-and-effect relationship between an individual training program and organizational performance can be difficult given the multitude of variables that may affect organizational performance (Kirkpatrick & Kirkpatrick, 2012; Phillips & Aaron, 2006). Consequently, the data collection and analysis required to evaluate Levels III (behavior) and IV (results) become more rigorous.

Taking a holistic approach to determine the overall impact to organizational performance by considering soldiers' reactions to the training, level of knowledge gained, application of knowledge and skills back to the organization, and organizational impact provides feedback that is rich in context, multilevel, and provides corroborative evidence to the findings (Griffin, 2012).

The training of individual soldiers in the Army's institutional environment, euphemistically referred to as "the schoolhouse," presents challenges for anyone attempting to assess the impact of that individual training on larger organizational performance. Soldiers come to the institution either as part of their initial entry training (e.g. basic combat training or officer basic branch training), or to attend a professional military education course or special skill development course. The training courses that these soldiers attend include soldiers going to or coming from many units across the Army, and the operational performance of those many units is the sum of the individual performance of each soldier in the unit. This creates two significant challenges for evaluation. The first challenge is tracking these soldiers to their units and then assessing the degree to which they are able to apply what they learned in the course. The second challenge is establishing a means to correlate the individual knowledge and skills the soldiers developed in the course with the unit's operational performance.

The DCT-MT course presented a unique opportunity to overcome both of these challenges. Each iteration of the course was delivered to soldiers exclusively from the same brigade, and these brigades then deployed to Afghanistan at some point after their soldiers had received the DCT-MT training. Additionally, the means to collect data related to IED incidents was already in place, providing a mechanism for evaluating unit operational performance.

The Dismounted C-IED Tactics Master Trainer (DCT-MT) course evaluation plan was developed to evaluate the effectiveness of the DCT-MT course and to determine the impact the course had on trained units' operational performance in theater compared to baseline units that did not receive the training. The Kirkpatrick's four-level evaluation model was used for the evaluation. The evaluation employed a mixed methods design to evaluate Levels I and II during the training and Levels III and IV during and after the unit's deployment.

METHODOLOGY

The Kirkpatrick's four-level evaluation model was selected as the framework because it allowed the evaluator to adapt the model to the needs and requirements of evaluating an institutional training course. The flexibility and adaptability of the model allowed the criteria of the evaluation to be aligned with the indicators, or measures of success, for each evaluation level. The four-level model has been the gold standard for evaluating organizational training and provides a guide for the kinds of questions that should be asked and the criteria for each level (Bates, 2004). The accuracy to the measure of effectiveness of the evaluation is increased because of the successive nature of the model. Each level of the model is successive in that each level builds upon the previous level (Tan & Newman, 2013). With the progress of each level, the evaluation process becomes more laborious, time consuming, and more difficult to measure.

Level I evaluation measured soldiers' reactions to the course. End of course surveys were collected at the end of each class and were analyzed to determine the soldiers' perceptions about the course and their recommendations for improvements. The survey consisted of open ended questions asking soldiers to recommend improvements to the training. Soldiers also rated the different aspects of the training and their level of preparedness to perform duties as the DCT master trainer. A 5-point Likert scale was used for the ratings.

Level II evaluations measured soldiers learning. Pre and posttest scores were compared to determine if and how much the soldiers' knowledge increased at the end of the course. A total of five tests were administered throughout the course to measure the increase in soldiers' knowledge. The pre and posttests consisted of 30 questions that contained multiple choice and matching questions. The test questions for the pre and posttests were the same but not presented in the same order. Tests 1, 2, and 3 each had 25 questions and consisted of multiple choice and true/false questions. Soldiers were required to score 80% or better on each test. Soldiers who scored below the required 80% were retested using another version of the same test that consisted of the same number and types of questions.

Level III evaluation measured the application of knowledge to determine if soldiers applied what they learned during the training once they returned to their units. A mixed method design was used because by itself, a

quantitative or qualitative method could not provide the answers to the evaluation questions that were used to guide this evaluation plan. Mixed method allowed for an in depth analysis and an opportunity to answer the “why’s” and “so what” to certain questions. The technique used to collect data for evaluating Level III was methodological triangulation. Methodological triangulation involved collecting data from multiple sources (e.g. from focus groups, surveys, or interviews). The results from each of these sources were compared to determine if the results were the same or similar (Guion, Diehl, & McDonald, 2011). The evaluation team anticipated the challenges of collecting data from theater and did not rely on one or two sources of data. Redundancy in the data collection instruments and methods ensured that the team would gather sufficient data to identify emerging trends.

Surveys and observations were collected from multiple perspectives to measure the application of knowledge to determine if soldiers applied what they learned during the training back in their units. Surveys were collected from soldiers and their respective commanders while they were deployed and conducting operations in Afghanistan. Focus group sessions were conducted after the units returned from Afghanistan with the graduates of the course, and when possible, their company commanders. The Soldier Survey, which had been routinely administered by the Maneuver Center of Excellence (MCoE) for units returning from combat deployments, provided a unique opportunity for the evaluation team to gain insights about the performance of the master trainers from the perspectives of soldiers and leaders. The evaluation team developed additional questions that were included in the survey. Observations by the Joint Expeditionary Team (JET) from theater were used to determine the application of knowledge. The JET advisors conducted C-IED Advisory Missions (CAMs) across the battle-space in Afghanistan where they were embedded in units. During these missions, the advisors observed the master trainers within their units. The advisors completed a CAM summary report after each of their missions and these reports were provided to the evaluation team.

Level IV evaluation measured units’ operational performance in combat. The number of casualties caused by IEDs, IEDs that were found and cleared, and IED detonations were the variables used in this study. This data was already being collected by the Joint Trauma Analysis and Prevention of Injury in Combat (JTAPIC) team. The evaluation team submitted a request for information (RFI) to JTAPIC. The JTAPIC managed all RFIs and the analyst from the MCoE Dismounted Incident Analysis Team (DIAT) completed the analysis to answer the RFIs. The DIAT analyst flagged all casualty producing events or activities for the unit from the Combined Information Data Network Exchange. The data collected were vetted to determine if the event occurred during a mounted or dismounted operation and what type of weapon system caused the injuries. The DIAT analyst pulled data from the Operations Research System Analysis’ database to analyze the number of found and cleared IEDs and the number of IED detonations. It was not possible to collect data from Afghanistan for all units that completed the course due of the length of the units’ deployments and the time line for this evaluation. RFIs were submitted for four units that received the DCT-MT course training and for the units that they replaced in Afghanistan, who had not received the DCT-MT course training. Detailed reports for each of these units were completed by the DIAT analyst and were used in the analysis to determine the overall impact of the course on the operational performance of these units.

RESULTS

Each level of the evaluation was treated as an independent evaluation with its own data collection instruments and plan. Multiple data collection instruments were used throughout the evaluation and data collection process, to include four surveys for commanders, four surveys for soldiers, interviews for both soldiers and commanders, and five tests that the soldiers completed during the training. Over 30,000 data points were collected through the surveys, which did not include the data points collected by the Testing and Evaluation Team’s Soldier Survey.

There were 21 classes conducted for a total of 511 soldiers and 14 civilians and contractors who completed the training. Classes were conducted for 19 brigade combat teams, one for the operations group at the National Training Center, and one class for a group of soldiers from different units at Fort Benning. Data collection for Level I began during the eighth class of fiscal year 2013 and Levels II through IV began with the first class in August 2012. All data collection concluded with the eighteenth class in FY 2013.

Level I: Reaction

Level I evaluation measured soldiers reaction to the training, course materials, and the events throughout the course. Soldiers' reactions were collected from the end of course survey completed at the end of training. The purpose of the survey was to determine how soldiers perceived the training, relevancy of the training, and recommendations for improving the course. Soldiers rated nine aspects of the course on a 5-point Likert scale that included the following response options: 1-Very Poor, 2-Poor, 3- Average, 4-Good, and 5-Very Good.

The measure of success established for Level I evaluation was an average rating of 4 or higher for each of the nine rated aspect of the course. These included the overall conduct of course, professionalism of the cadre, relevance to combat operations, facility and equipment, safety during training, quality of instruction, quality of training material, quality of training areas, and class schedule. The *n* for all the rated aspects was 283 except for quality of training material and class scheduling, which were 282 and 280. The ratings for each aspect of the course were above the desired average of 4. The soldiers positively perceived the training and considered the mobile training team (MTT) instructors to be professional and knowledgeable. Figure 1

summarizes the average rating for each rated aspect of the course.

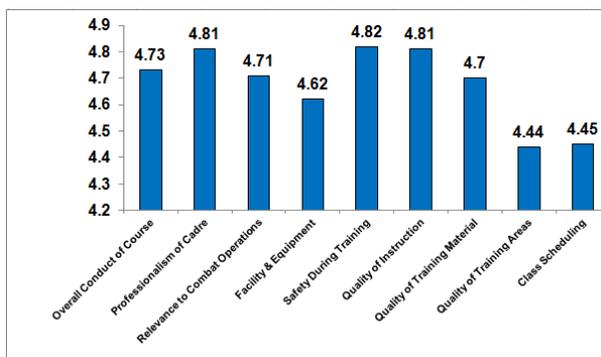


Figure 1. Level I: Reaction Rating

Level II: Learning

The most direct way to measure learning in a training environment is by administering pre and posttests (Praslova, 2010). The measure of success established for this level of evaluation was an average increase of 25% or higher from the pre to posttest scores. The average pre-test score was 47% and the average post-test was 77%. The average increase from pre to posttest score was 31%. Of the 19 classes included in the analysis, two classes did not achieve the desired average increase of 25% or higher. Four classes had an average posttest score of 80% or higher and an average increase of more than 30%. The frequency distribution of the scores is presented in Table 1.

A one sample *t* test revealed a statistical significance between a soldier's rank and posttest score ($p=.000$)

and the ANOVA revealed $p=.000$. Correlation analysis of ranks and posttest score revealed $r=.231$ at $p=.000$.

		Statistics		
		Pretest	Posttest	Average Increase
N	Valid	484	484	484
	Missing	0	0	0
Mean		46.06	77.44	31.47
Median		47.00	80.00	32.00
Mode		40	80	32
Std. Deviation		12.690	10.179	13.584
Minimum		12	48	-4
Maximum		96	100	72

Table 1. Test Scores Frequency Distribution

Level III: Behavior

To evaluate the application of knowledge that soldiers gained from the training back to their unit, data had to be collected while the units were deployed to Afghanistan. There are several inherent challenges with collecting application data from theater. These challenges included limitations on how data could be collected, potential low response rate for emailed surveys, and competing for soldiers' time with other operational requirements (e.g. completing a survey versus preparing for a mission). Additionally, computers, phones, and Internet were not available at every forward operating base or combat outpost. To account for these challenges, four different data collection instruments were developed to gather data for Level III analysis. Table 2 depicts a summary of the data collected from the four different sources.

Instruments	Class No.	n	Results
In theater Follow-on Survey	1-12	<ul style="list-style-type: none"> • 8 Soldiers • 10 Commanders 	Soldiers performed their duties as master trainers when dismounted patrols were a part of unit's mission, had the support of unit leaderships, and soldiers duty position allowed them to train soldiers. Soldiers applied their knowledge and assisted units with updating unit standard operating procedures (SOP) and advised units on training.
	2-12	<ul style="list-style-type: none"> • 7 Soldiers • 3 Commanders 	
	3-13	<ul style="list-style-type: none"> • 18 Soldiers • 6 Commanders 	
Focus Group	2-13	<ul style="list-style-type: none"> • 12 Soldiers • 3 Commanders 	Soldiers recommended that commanders and senior noncommissioned officers (NCO) in units attend training to gain a better understanding of their capabilities as master trainers. Although not able to develop or implement a formal training plan, master trainers assisted units during rehearsals prior to each mission and conducted hip pocket and refresher training. Commanders trusted the recommendations on the integrating of handheld devices into patrols from their master trainers.
	3-13	<ul style="list-style-type: none"> • 21 Soldiers 	
	4-13	<ul style="list-style-type: none"> • 5 Soldiers 	
JET CAMSUM	3 BDEs	NA	JET advisors observed noticeable differences between trained and untrained units. Trained units had a better understanding of the capabilities of the equipment and observed NCOs conducting training on the devices for their soldiers.
Soldier Survey	3 BDEs	<ul style="list-style-type: none"> • 23 Leaders 	85% of leaders were satisfied or very satisfied with the performance of master trainer. 91% of the leaders were advised on the employment of C-IED handheld devices. 58% of leaders received a training plan from their master trainer.
	1 MCCC	<ul style="list-style-type: none"> • 10 Leaders 	

Table 2. Level III Summary

The themes or trends that emerged from each of the data collection instruments were compared to each other to determine if there were any commonalities. Themes that emerged from these sources included:

- Most soldiers did not have an opportunity to apply their learning to conduct training for their units prior to their deployment or during their combat training center (CTC) rotation because they received training shortly after their unit's CTC rotation or with less than 45 days before deploying. Units were focused on packing and block leave and did not have access to handheld devices to conduct training.
- Once in theater, soldiers did apply their knowledge. However, units that were mounted (i.e. conducted most of their operations using vehicles) did not find the course useful and the master trainers did not apply the knowledge they learned from the course.
- Although units did not implement a full training plan, the master trainers assisted and trained soldiers in units during rehearsals prior to missions.
- Soldiers applied their knowledge if certain conditions were met. These conditions included:
 - Leadership understood the capabilities of the master trainer and the purpose of the training.
 - Soldiers who attended the course met the minimum rank recommended for this course; several soldiers stated that unit leadership did not listen to them because they were Specialists.
 - Soldier's duty position allowed them to apply their learning (e.g. the soldier was a platoon sergeant)

and conducted training for his platoon).

- Factors that negatively affected soldier’s application of knowledge:
 - Unit leadership’s lack of understanding of the capabilities of the master trainer.
 - Soldiers were sent to training to fill seats.
 - Units not knowing that they had a DCT-MT graduate within unit or graduate did not deploy with the unit.
 - Unit conducted primarily mounted operations.
 - Soldier’s duty position did not allow them time to develop training plans or train units.

Level IV: Results

Level IV evaluation measured the impact the training had on the organization’s operational performance. The three variables used to measure the operational performance of the trained units were the number of casualties produced by IEDs, number of IEDs found and cleared, and the number of IED detonations. A one-to-one comparison of these variables was conducted between one unit that participated in the DCT-MT course (trained unit) and one unit that did not participate in the DCT-MT course (untrained unit). The trained unit in each of the comparisons replaced the untrained unit in the same area of operation. The time span of this evaluation, length of deployment, and the shift to retrograde operations affected the availability of data for Level IV analysis. Data was available for only four separate comparisons to be conducted.

Table 3 depicts the actual numbers for the variables in the analysis between trained and untrained units.

Variables	Comparison 1				Comparison 2				Comparison 3				Comparison 4			
	Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit	
	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO
Casualty	41	200	24	34	83	155	11	70	13	13	45	98	7	12	3	13
Found/Cleared	389	1234	37	237	196	417	135	1011	112	161	37	410	30	234	15	168
Detonation	29	131	21	165	134	284	53	181	89	400	73	436	28	182	11	205

Table 3. Comparison Summary

During the data analysis for the three variables (casualty, found/cleared, and detonation), confounding factors (e.g. seasons, retrograde operations, units’ locations, and missions) were considered. A direct comparison of the actual numbers for each variable was not conducted. The ratio of each variable between the untrained and trained units and all units within the area of operations were used for the analysis to compare these variables between the untrained and trained units. Ratios, rather than actual counts of each variable, were used to account for enemy activities, location of units, and the number of IEDs within each regional command.

The percentage calculated for each variable was a ratio of the unit’s actual numbers for each variable and the total number from all units within the area of operation. The casualty ratio for all units were calculated by dividing the total number of casualties produced by IEDs for the unit by the total number of casualties produced by IEDs for all units throughout the area of operations during the same time frame. The total found/cleared ratio was calculated by dividing the total number of IEDs the unit found/cleared by the total of IEDs found/cleared by all units in the AO. The detonation ratio for each unit was calculated by dividing the unit’s total number of IED detonations by the total IED detonations by all units in the AO. The results of the comparisons for Level IV are summarized in Table 4.

Variables	Comparison 1		Comparison 2		Comparison 3		Comparison 4	
	Untrained Unit	Trained Unit						
Casualty (%)	21%	71%	54%	16%	100%	46%	58%	23%
Found/Cleared (%)	32%	16%	47%	13%	69%	9%	13%	9%
Detonation (%)	22%	13%	47%	29%	22%	17%	15%	5%

Table 4. Level IV Comparison

Successful operational performance was measured by a decrease in all three ratios for the trained units in comparison to the ratios of untrained units. The rationale for these measures is that the soldiers were trained in the DCT-MT course to mitigate risks, thus reducing the number of injuries sustained from IEDs; avoid vulnerable points and areas, which should decrease the IED detonations, and an increase in ground sign awareness should decrease the number of found/cleared IEDs. Figure 2 depicts the ratio comparison of each variable and each ratio is discussed in the following sections.

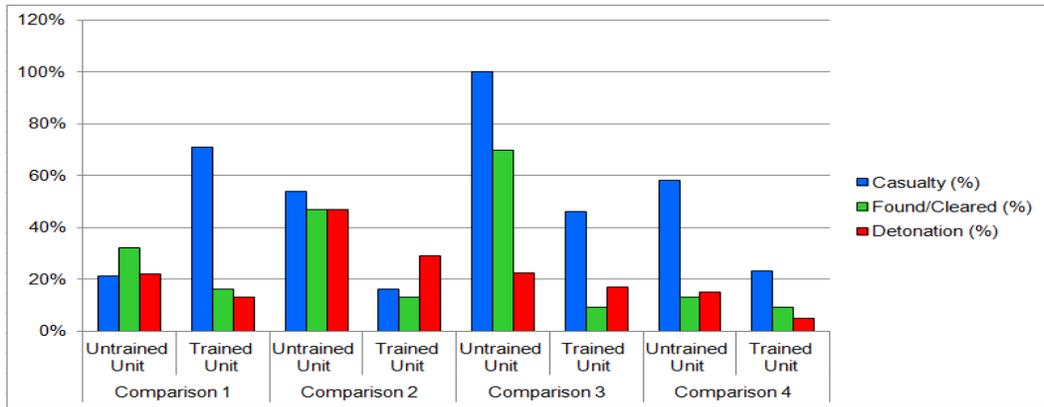


Figure 2. Level IV Comparison

Casualty Ratio

The actual number of casualties and the ratio for each comparison are included in Table 5.

Casualty	Comparison 1				Comparison 2				Comparison 3				Comparison 4			
	Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit	
	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO
Actual No.	41	200	24	34	83	155	11	70	13	13	45	98	7	12	3	13
Ratio	21%		71%		54%		16%		100%		46%		58%		23%	

Table 5. Casualty Ratio Comparison

In comparison 1, the trained unit suffered a catastrophic event and sustained 13 casualties in one incident. The untrained units in comparisons 2, 3, and 4 had a casualty ratio of 54% or higher. In comparison 3, the untrained unit was the only unit in the AO that had casualties during that time frame.

The sample *t* test revealed that there was no statistical significance ($p=.135$) between the number of casualties and the DCT-MT course. The correlation between the number of casualties and the DCT-MT course was $r=-.323$ but was not statistically significant ($p=.223$).

Found and Cleared Ratio

The actual number of IEDs found and cleared and the ratio are included in Table 6.

Found/Cleared	Comparison 1				Comparison 2				Comparison 3				Comparison 4			
	Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit	
	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO	Unit	AO
Actual No.	389	1234	37	237	196	417	135	1011	112	161	37	410	30	234	15	168
Ratio	32%		16%		47%		13%		69%		9%		13%		9%	

Table 6. Found/Cleared Ratio Comparison

The found/cleared ratio was lower for all trained units in all four comparisons. A sample *t* test revealed a statistical significance between the DCT-MT course and the number of found/cleared IEDs ($p = .036$). At the 99% confidence level, the analysis of variance revealed the training resulted in a lower number of found/cleared IEDs ($p = .001$).

Correlation analysis of the number of IEDs found/cleared and the DCT-MT course revealed $r = .728$ at $p = .001$. The number of found/cleared IEDs had a positive correlation to the number of casualties with $r = .612$ and $p = .011$.

Detonation Ratio

The actual number of IED detonations and ratio for each comparison are included in Table 7.

Detonation	Comparison 1				Comparison 2				Comparison 3				Comparison 4			
	Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit		Untrained Unit		Trained Unit	
	Unit	AO	Unit	AO												
Actual No.	29	131	21	165	134	284	53	181	89	400	73	436	28	182	11	205
Ratio	22%		13%		47%		29%		22%		17%		15%		5%	

Table 7. Detonation Ratio Comparison

Detonation ratios for all trained units were lower than for all untrained units. A sample t test revealed a statistical significance between DCT-MT course and the number of IED detonations ($p = .006$). At the 95% confidence level, the analysis of variance revealed the training resulted in a lower number of detonations ($p = .048$).

Correlation analysis of the number of detonations and the DCT-MT course revealed $r = .502$ at $p = .048$. The number of IED detonations was positively correlated to the number of casualties with $r = .757$ at $p = .001$. A unit with reduced detonations would also have a reduced number of casualties. The number of detonations was positively correlated to the number of found/cleared IEDS with $r = .830$ at $p < .001$.

DISCUSSION

The measure of success for Level I evaluation was a rating of 4 or higher on the Likert scale for each rated aspect of the course. Each rated aspect received a rating well above 4. Soldiers were satisfied with the quality of the instruction and the professionalism and knowledge of the instructors. The measure of success for Level II evaluation was an average increase of 25% or higher from the pre to posttest scores. Of the 19 classes, classes 2-12 and 3-12 did not meet this standard as both had an average increase of 21%. Although soldiers from both classes perceived the training positively, they missed some training because their units were preparing to deploy.

Classes 2-12 and 3-12 rated seven of the nine rated aspects of the course higher than class 10-13 did, which had the highest increase between pre and posttest scores. Classes 2-12 and 3-12 also rated all nine aspects of the course higher than all the other classes in the analysis. The two rated aspects that received the lowest rating from classes 2-12 and 3-12 were the facility and equipment and class scheduling, respectively 4.63 and 4.62. The two highest ratings were the quality of the instruction and professionalism of the cadre, respectively 4.97 and 4.91. The 21% increase from pre and posttest scores for classes 2-12 and 3-12 indicated an increase in knowledge and learning occurred during the training. However, a positive reaction to the training did not guarantee a significant increase in learning. For this evaluation, a significant increase in learning was a 25% or higher increase between pre and posttest scores. Other factors had an impact on learning. For classes 2-12 and 3-12, these factors included distractions from unit requirements and missing blocks of instructions. These factors impacted learning more than the positive reaction to the training.

The themes that emerged from the four different data sources for Level III evaluation were compared to determine if the conclusions from each of the sources were similar or the same. The conclusions were similar and trends for successful application of knowledge emerged. Master trainers in units that conducted dismounted patrols used the new skills and knowledge that they acquired from the training to assist their units in conducting rehearsals, refresher training, or developing unit SOPs. In addition to using their knowledge and skills, the JET advisors observed a positive change in the behaviors of the master trainers compared to other NCOs from units that did not receive the training.

Although classes 2-12 and 3-12 did not meet the benchmark of 25% increase between pre and posttest scores, the 21% increase in score indicated learning did take place. While in theater, classes 2-12 and 3-12 completed the in theater follow on survey as part of the Level III evaluation. The soldiers whose units conducted dismounted patrols and whose duty positions allowed them to perform their duties as a master trainer indicated that they

applied the new skills and knowledge they learned from training. These soldiers conducted training for their units, advised unit leadership, and assisted in the development and refinement of unit standard operating procedures. The soldiers of these two classes had a positive reaction to the training but collectively did not meet the 25% increase in test scores. They did, however, apply what they learned from the training in performing their duties as the master trainer for their units.

Measuring the effectiveness of the training on the unit’s operational performance is not as definitive as in Level I or Level II. Although three of four trained units had a lower number of casualties and all four trained units had lower numbers of IEDs found/cleared and detonated IEDs, other variables affected these numbers. Efforts were taken to account for confounding variables but not all confounding variables could be accounted for during data analysis for Level IV evaluation, such as catastrophic events and enemy activities. As illustrated in the casualty ratio of the trained unit in comparison one, a catastrophic event that resulted in 13 casualties had a significant impact on the casualty ratio. Further analysis was conducted to determine if catastrophic events affected other units in the same manner. For all units, both trained and untrained, the highest number of casualty per a single event was excluded from the unit and the AO’s total number. The casualty ratios between trained and untrained units were recalculated. The comparison is presented in Table 8.

Casualty Ratio	Comparison 1		Comparison 2		Comparison 3		Comparison 4	
	Untrained Unit	Trained Unit						
With Event	21%	71%	54%	16%	100%	46%	58%	23%
Without Event	18%	52%	50%	11%	100%	38%	44%	17%

Table 8. Catastrophic Event Analysis

Even with the recalculation of the casualty ratio, the trained unit’s casualty ratio in comparison 1 is still higher than the untrained unit. The trained unit’s ratio in comparison 2, 3, and 4 are still lower than the untrained unit’s casualty ratio.

Another confounding variable that had to be accounted for during data analysis was enemy activity. In comparison 3 in Level IV analysis, the trained unit had more casualties from indirect fires (IDF) than the untrained unit. The number of casualties from IDF for the untrained unit was 54% of all the IDF within the AO. The number of casualties for the trained unit from IDF was 79% of all the IDF within the AO.

Class 1-12 was the only class that data was collected for all levels of the evaluation. At each level of the evaluation, class 1-12 is the ideal class and the results of each level of the evaluation exceeded the minimum standard for each measure of success. For Level I, class 1-12 rated all aspects of the course higher than all other classes. The reactions to the training were positive, with the lowest rating of 4.86 for the professionalism of the cadres and the highest rating of 4.93 for the relevancy to combat operations. For all other classes, these ratings were 4.81 for the professionalism of the cadres and the highest rating of 4.71 for the relevancy to combat operations. Class 1-12 had the second highest posttest score average of 86% and an increase of 31%. The in theater follow on survey completed by the soldiers and company commanders of class 1-12 for Level III analysis indicated that soldiers applied their knowledge and skills. They assisted in the development of unit SOPs, conducted unit level C-IED certification training, conducted rehearsals prior to patrols, and advised their leaders.

When the master trainers performed their duties in advising leaders, integrating handheld devices, and planning tactical employment of handheld devices to counter-IED related threats targeting dismounted patrols, then the number of casualties, found and cleared IEDs, and IED detonations of trained units should be less than those of untrained units. Of the four comparison analysis conducted for Level IV evaluation, class 1-12 performed as expected in each of the categories of analysis, without having to account for confounding variables.

Several factors contributed to the overall success of class 1-12. The leaders of the unit sent the soldiers with the targeted rank and Military Occupation Specialties (MOS). Every soldier in the class was a Sergeant or above. All soldiers were from the targeted MOSs (Infantrymen, Combat Engineers, and Cavalry Scouts). The soldiers reacted positively to the training and perceived the training to have relevancy to combat operations. The unit leadership

utilized the master trainers and allowed them to perform their duties by applying their knowledge and skills gained from the training. The soldiers also had duty positions, such as a platoon sergeant or a squad leader, which gave them direct influence over their soldiers' training or rehearsals prior to each patrol.

Limitations

This evaluation was limited by the length of time it took to collect data for Level I to IV, which was approximately 14 months. The time frame encompassed the time a unit completed the training to deploying to conducting focus groups during post deployment data collection sessions. Data collection could not be collected at each stage of the evaluation process, which hindered the ability to conduct a complete analysis of Levels I through IV for more than one class. Class 1-12 was the only class of the 19 that data for Levels I through IV was collected.

The ability to generalize the findings of this evaluation is limited by the low response rates for the surveys. Although the findings of the responses from the in-theater follow on surveys from soldiers and company commanders was used for Level III analysis as a source of data, the findings from the surveys could not be generalized to the larger population because of the small sample size and low response rates.

The findings and recommendations provided in this evaluation are limited to the interpretation of the data by the evaluator. The evaluator has subjective biases and professional and military experiences that could affect how the qualitative data was analyzed, coded, and interpreted. In the same manner that methodological triangulation was used for Level III evaluation, investigator triangulation could be used in the future, which involves other evaluators during the data analysis phase of the evaluation. Investigator triangulation can increase the validity of the findings if the evaluators arrive at the same or similar conclusions.

Recommendations

The four-level evaluation model is designed for evaluating training. An evaluation does not have to use all four levels of evaluation. The evaluation purpose, the significance of the training program, how the evaluation will be used, the resources available to conduct the evaluation will all impact the decision of which levels of evaluation should be conducted. The four-level model used in this evaluation project provides a framework to follow and streamlines the process of answering the evaluation questions. The model is adaptive and can be tailored to meet the needs of the organization. As such, this model should continue to be used for future training evaluation projects.

One of the challenges in conducting training program evaluation is accounting for the many variables other than the training intervention that may affect the outcome. An issue encountered in this evaluation was the lack of a controlled environment. Creating an ideal controlled experiment is unrealistic (Sackett & Mullen, 1993). Therefore future evaluation efforts must identify beforehand the many uncontrollable variables and establish rigorous means to account for those variables. Evaluation projects in the future could use combat training center rotations and home station mission readiness exercises to evaluate a unit's performance in a controlled environment. Rather than comparing the ratios of the units to other units in the area of operation, actual performance data from units could be analyzed.

When evaluating Level III or higher of other training programs, methodological and investigator triangulation should be employed, especially when conducting qualitative data analysis. Having more than one analyst analyzing the same data could provide deeper insights into the understanding of emerging themes and increase the level of confidence in the data. The interpretation of the data is also not solely dependent on one analyst. This evaluation process combined the different knowledge and capabilities of several different entities. Each group directly contributed to the data analysis and collection process and the overall success of the project. The collaborative effort throughout this process provided different data collection instruments, continuity in data analysis, and filled several capability gaps within the design and evaluation team. The collaborative effort also enhanced both the quantitative and qualitative data collection. The collaborative effort should not only be replicated but continuously sustained, especially on future evaluation projects that are as rigorous as this one.

CONCLUSION

Based on a need to demonstrate the utility of a new Army training course, a detailed evaluation plan was developed and implemented for the DCT-MT course. This evaluation plan applied the Kirkpatrick four-level framework, evaluating the learners' reaction to the training, their learning, the application of what they learned in their unit, and the impact that this training course had on units' operational performance while conducting combat operations in Afghanistan. The results of this evaluation indicated that units with soldiers that completed the DCT-MT course were more effective in countering the IED threat than units that did not have soldiers complete the course. While these results are not conclusive, they do provide valuable information to stakeholders in determining the overall value of this particular training course.

More importantly, this evaluation project provided valuable lessons for the evaluation team to apply in future training program evaluations. The methods used for this evaluation will be modified as appropriate and applied to the evaluation of other training courses and programs. Many lessons were learned by the evaluation in the course of this evaluation project. Those lessons will be applied to future evaluation projects.

REFERENCES

- Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning*, 2, 341-347.
- Bober, C. & Bartlett, K. (2004). The utilization of training program evaluation in corporate universities. *Human Resource Development Quarterly*, 15(4), 363-383.
- Griffin, R. (2012). A practitioner friendly and scientifically robust training evaluation approach. *Journal of Workplace Learning*, 24(6), 393-402.
- Guion, L.A., Diehl, D.C., & McDonald, D. (2011). Triangulation: Establishing the validity of qualitative studies. Retrieved from: <https://edis.ifas.ufl.edu/fy394>
- Kirkpatrick, J. D. & Kirkpatrick, W. K. (2012). The Feds lead the way in making training evaluations more effective. *T + D*, 66(9) p. 56-61.
- Patel, L. (2010). Overcoming barriers and valuing evaluation. *T & D, February 2010*, 62-63.
- Phillips, J. J. & Aaron, B. C. (2008). Isolation of results: Defining the impact of the program. In J. J. Phillips & P. P. Phillips (Eds.), *ROI Fundamentals: The measurement & evaluation series (Vol. 3)*. San Francisco: Pfeiffer.
- Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Education Assessment Evaluation and Accountability*, 22, 215-225.
- Sackett, P. R., & Mullen, E. J. 1993. Beyond formal experimental design: Towards an expanded view of the training evaluation process. *Personnel Psychology*, 46, 613-627.
- Steensma, H. & Groeneveld, K. (2010). Evaluating a training using the "four levels model." *Journal of Workplace Learning*, 22(5), 319-331.
- Tan, K. & Newman, E. (2013). The evaluation of sales force training in retail organizations: A test of Kirkpatrick's four-level model. *International Journal of Management*, 30(2), 692-701.