Evaluating and Benchmarking Non-Governmental Training Programs: An Analytic Hierarchy Approach

Salah R. Agha *

Industrial Engineering Department, Faculty of Engineering, Islamic University-Gaza

Abstract

This study uses Analytic Hierarchy Process (AHP) as a multi-criteria decision making technique for evaluating training programs. Critical factors characterizing successful training programs are first identified using interviews with experts along with questionnaires. Once the factors were identified, the hierarchy was constructed and the factors were ranked according to their importance with respect to achieving the overall goal set for training. Results show that, of these critical factors, content and trainer rank the highest, while logistics came last. Then, two existing non-governmental training programs are selected, evaluated, and compared. Moreover, and as requested by one of the administrations of one of the evaluated programs, a benchmarking alternative is derived based on the results of the two training programs. Finally, recommendations are given to each of these programs in particular and training programs in general, for possible areas of improvements based on the strengths and weaknesses of each program as compared to the benchmark.

1. Introduction

The establishment of the Palestinian National Authority and redeployment of the Israeli forces in Gaza Strip had put more pressure on the Palestinians to start their own businesses and improve their economy. Donors and Palestinian Authority have recognized the need for different training programs in the Palestinian territories to help develop the Palestinian economy. Figure 1 shows the different training programs in Gaza Strip. These training programs include governmental, international and non-governmental (NG) training programs. As for the governmental programs, they include the programs that are offered and administered by the different ministries. These training programs are normally directed towards the employees of the specific ministries. While the international training programs are the ones that are internationally funded and managed. The target groups of this type of training programs include employees and fresh graduates. Finally, the NG programs are the ones that are funded by international donors and managed by local institutions and universities and they constitute a large proportion of the existing training programs. In general, most existing training programs can be classified to management and professional training programs. Examples of management training programs include but -

not -limited to time management, proposal writing, etc.…, while the professional training programs may include Microsoft Certified System Engineer (MCSE), AutoCAD, Oracle, Primavera, EPANET, Geographic Information Systems (GIS), Cisco Certified Network Associate (CCNA), International Computer Driving License(ICDL), Statistical Package for Social Sciences (SPSS), 3DMax and others. Due to fierce competition and relative ease of data collection, this paper focuses on evaluating this type of programs using the Analytic hierarchy Process (AHP).
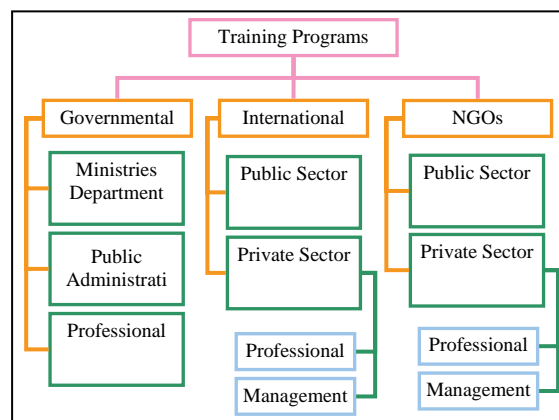

Figure 1: Types of Training Programs in Gaza Strip [1]

---

* Corresponding author. e-mail: aghasr@yahoo.com

AHP is a decision making tool that allows the decision maker to model a complex problem in a hierarchical structure showing the relationships between goal, criteria, sub-criteria, and alternatives. A multi-criteria decision making methodology allows subjective as well as objective criteria to be considered in the evaluation process. AHP has a variety of applications in different fields, such as planning, selecting, evaluating, and benefit/cost analysis. These fields are found in different arenas including personal, social, manufacturing, political, engineering, education, industry, government, sports, and management [2-13].

The selection of AHP in this research is because AHP provides a realistic description of the problem by incorporating all aspects in the hierarchy. Moreover, AHP provides a useful mechanism for checking consistency of the evaluation measures and thus reducing bias in decision-making. In addition, AHP may help in SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis by identifying strengths and weaknesses of training programs and paves the way for a successful benchmarking.

This study is motivated by the fact that in Gaza Strip, and after the Israelis redeployment, training programs spread widely in different fields for different categories of trainees. The purpose of most of these training programs is to bridge the gap between the current practices and the state of the art. The existing evaluation processes lack scientific basis. In other words, most of the research in this area is merely of a data collection type with no in depth analysis. According to Shaaban [14], the weaknesses of the training programs are not a matter of lack of resources; rather they are a management-related problems stemming from the fact that training programs impact is not systematically evaluated. Thus, the purpose of this paper is to evaluate these training programs and recommend necessary actions for improvement using the AHP methodology as a tool. This paper is organized as follows: section two reviews the methods used in training programs evaluation, followed by AHP methodology in section three. In section four, the specifics of the application along with the data collection methodology are given. Results and analysis are given in section five followed by conclusions and suggestions for these programs and others in section six.

1.1. Training Programs Evaluation

The process of evaluating training programs consisted of several and sequential steps. If these steps were properly performed, the evaluation results would certainly be helpful and lead to making sound decisions regarding improvement efforts. The evaluation process started with collecting the needed data using the suitable data collection methods such as questionnaire, interviews, documentation review and observation, then analyzing and interpreting data. In the analysis step, it was necessary to start with the evaluation goals in order to organize data and focus the analysis. For example, if the goal was to improve a training program by identifying its strengths and weaknesses, data could be organized into program strengths, weaknesses, and suggestions to improve the program. Finally, a suitable model for evaluating training programs, such as Kirkpatrick's four levels, Quality Function Deployment (QFD), or Analytic Hierarchy Process (AHP) was selected.

2. AHP Methodology

AHP is a systematic procedure that organizes the basic rational of the decision problem by breaking it down into smaller parts, then calling for a simple comparison with respect to pairs of judgments to develop priorities within each level of hierarchy. Finally, results are synthesized to obtain overall weights of the alternatives. The following paragraphs briefly describe the steps involved in AHP.

1. Breaking down the decision problem into a hierarchy of interrelated decision elements as shown in Fig. 2. This hierarchy consists of at least three levels, the goal of the decision problem is placed at the top, the second level includes the criteria affecting the decision, and the last level contains the alternatives, which are to be evaluated and compared. Additional sub-criteria levels may be added where needed.

2. Comparing the elements in each level in pairs using Saaty's scale, which is shown in Table 1.[15]. These comparisons are made using judgments based on knowledge and experience in accordance with their contribution to the main element in the level immediately above.
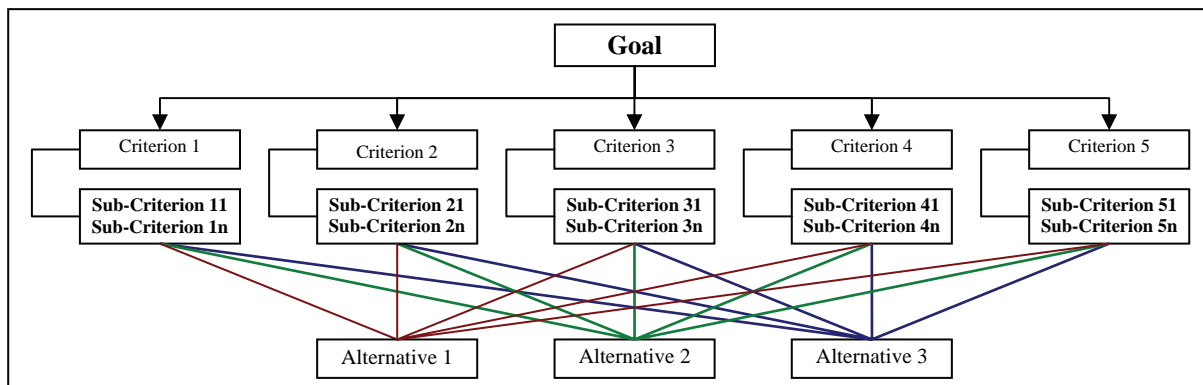

Figure 2: AHP Hierarchy

79

Table 1: Saaty's Scale of importance intensities

Intensity of Importance	Definition
1	Equal importance
3	Weak importance of one over another
5	Essential or strong importance
7	Demonstrated importance
9	Absolute importance
2, 4, 6, 8	Intermediate values between the two adjacent judgments

Due to reciprocity, the number of needed comparisons for (n) criteria is given by n*(n-1)/2

3. Calculating the average relative weight vector (the eigenvector).

4. Calculating the relative weights of the alternatives with respect to each criterion. For (n) criteria and (m) alternatives, the relative weights of the alternatives with respect to all criteria will form an m x n matrix.

5. Evaluating the consistency of the resulting weights
   Consistency is evaluated using the principal eigenvalues (λ_{max}) which is calculated through multiplying the pair-wise comparison matrix by the corresponding weights vector, then dividing the resultant matrix by the weights vector. Finally, the average value of the resultant vector λ_{max} is calculated. Once the value of λ_{max} is obtained, it is compared with the pair-wise comparison matrix size (n). If $\lambda_{max} = n$, a perfect consistency is said to exist, otherwise, there is an inconsistency with respect to the pair comparisons. Inconsistency is calculated using the consistency ratio (CR),

$$CR = \frac{CI}{RI} \qquad (1)$$

Where RI is a random number index, the values of which are shown in Table 2 [15]. While, CI is a random index of a randomly generated reciprocal matrix and it is calculated as

$$CI = \frac{\lambda_{max} - 1}{n - 1} \qquad (2)$$

Table 2: The reference values of RI

N	2	3	4	5	6	7	8	9	10
RI	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.51

If CR < 0.1, then with respect to the pair comparisons are said to be consistent, otherwise, reasons contributing to lack of inconsistency are investigated, and logic is used to revise the comparisons until CR is acceptable.

6. Calculating the overall weights of alternatives

The overall weights are determined by multiplying the relative weights of an alternative with respect to criteria by the relative weights of the corresponding criteria and summed over all criteria. Finally, sensitivity analysis shows to what extent the overall priorities are sensitive to changes in the importance of criteria. In other words, sensitivity analysis would answer different what if questions.

The more stable the ranking of the alternatives, the more confident management will be in the proposed choice. This analysis increases both the understanding of and confidence in the outcome of the AHP.

3. Application

Two of the prominent internationally funded non-governmental training programs are selected as an application in this study. Usually, these training programs are tailored according to the needs of the customers. These customers may be new university graduates and employees who are nominated by their companies to receive a given training to improve certain skills. Training sessions are performed at the trainers' site using their facilities. Though these training programs are externally funded, the trainees are sometimes asked to pay a nominal small amount of fees. Throughout this research and at the trainers' request, these training programs will be referred to as program A and B.

4. Data Collection

As mentioned earlier, the problem was divided into four levels, which were goal, criteria, sub-criteria, and alternatives. The elements of each level were identified through literature review and interviews with experts, while pair comparisons were obtained by interviewing experts and distributing questionnaires to the target group of trainees at both programs.

Four experts on the subject matter of training were carefully interviewed in order to obtain the importance of these criteria in achieving the goal of training. In other words, the discussion with the experts aimed at identifying the critical factors for successful training programs, making pair-wise comparisons among these factors using Saaty's scale, and identifying the elements of each factor. Figure 3. shows the criteria and the factors.

In order to elicit the specific data for this research, a questionnaire was designed. In its final form, the questionnaire has 37 closed questions that can be divided into three categories. The first category includes the respondent personal and demographic information. The second category questions evaluate trainees' satisfaction with the training programs they joined. This category includes the sub-criteria shown in Table 3. For clarity reasons, abbreviations were assigned to sub-criteria as shown in Table 3.

Table 3: Sub-criteria and Abbreviations

| Criteria	Sub-criteria	Abbreviation
Trainer	Trainer clearly and smoothly explains ideas	Idea
	Trainer gives enough time for participation of trainees	Participation
	Trainer covers sufficient material during training period	Sufficient Material
Content	Subjects (topics) of training are ordered from simple to complicated	Subject
	Handouts are clearly explained (charts, tables….)	Handouts
Time	T Span of training period is short (1,2, .., weeks)	Training Period
	Timing of sessions suits you	Timing of Session
	Number of training hours is sufficient	Training Hours
Logistic Facilities	Easy access to training place (Location)	Place
	Training place instills a sense of comfort (illumination, furniture, ventilation & calm…) (ergonomics)	Comfort
	There is hospitality during breaks	Hospitality
	Necessary equipment and augmentative tools is available (projector, computer…)	Equipments Availability
Fees	Fees are reasonable	Fees
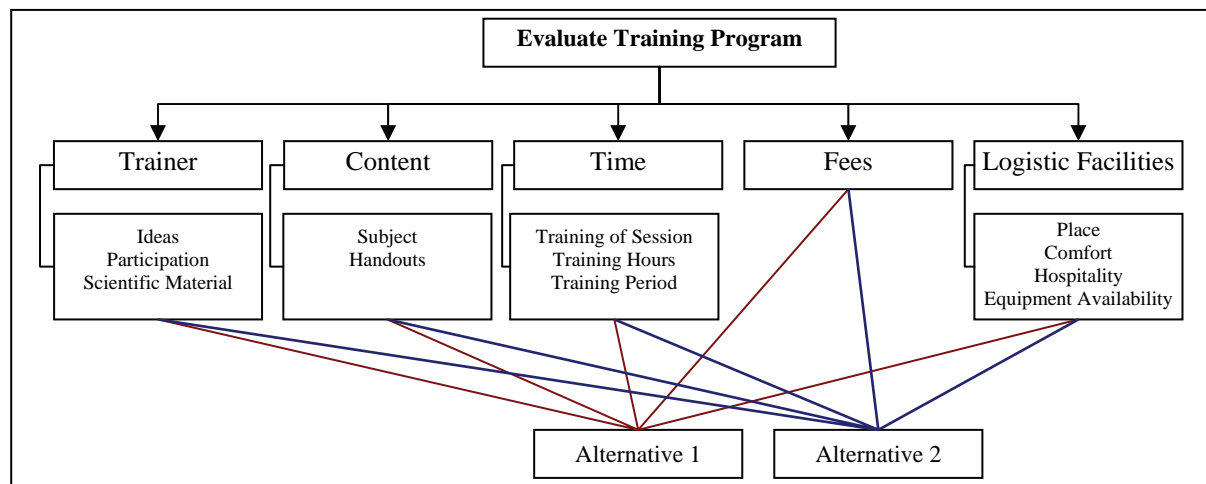
Figure 3: Training programs evaluation hierarchy with four levels

The final category determines the importance of the elements of each critical factor for successful training program. These elements are called sub-criteria and are presented in the third level of the hierarchy as shown in Figure 3.

As for the data needed for the fourth level, it was obtained through a questionnaire that was distributed to the target group of trainees who were chosen from the selected two training programs under study. The training programs provided a list of the trainees during the past year. The researcher contacted these persons and those who agreed to participate in the study were sent the questionnaire. The number of distributed questionnaires was 80. Seventy of them were received, but only 63 questionnaires were found valid. The questionnaire is considered invalid if it does not meet the acceptance criteria, which includes answering all questions, in addition to answering the test question right. The test question is a question that repeats the meaning of a specific question in the questionnaire but in a different format. Therefore, valid response rate was 78.75% as shown in Table 4.

Table 4: Valid response rate for the two training programs

TP	Distributed Questionnaires	Received Questionnaires	Valid Questionnaires	Valid Response Rate %
A	40	33	31	77.5
B	40	37	32	80
Total	80	70	63	78.75

Reliability and internal harmony were calculated for valid questionnaires. In this research, the Alpha-Kronbach test was used to measure the questionnaire reliability for both trainees' satisfaction category for each of the training programs A and B, and for the importance of the sub-criteria. After testing, it is found that all calculated values are within the acceptable range.

After the required data has been collected, Expert Choice (EC) along with MAT LAB software was used to implement the AHP (Expert Choice).

5. Results and Discussion

5.1. Criteria Results

Table 5. shows the average relative weights vector of each criterion with respect to the goal. It is seen from the table that content and trainer are the most important criteria as they represent 72.8%, whereas fees represent 14.5%, while time and logistic facilities rank the lowest among these criteria as they both represent 12.7%.

Table 5: Relative weights vectors of criteria with respect to the goal and their average

Criteria	Expert 1	Expert 2	Expert 3	Expert 4	Average Relative Weights Vector
Trainer	0.163	0.326	0.336	0.512	0.334
Content	0.499	0.380	0.430	0.267	0.394
Time	0.045	0.067	0.078	0.073	0.066
Fees	0.25	0.149	0.127	0.055	0.145
Logistic Facilities	0.043	0.077	0.029	0.093	0.061

5.2. Sub-criteria Results

Pair comparisons at the third level of the hierarchy were used to determine the relative weight of each sub-criterion with respect to its corresponding criterion. Results of relative weights for all sub-criteria are shown in Table 6. It is clear from Table 6. that idea, subject, timing of session, comfort, and equipment availability are the most important sub-criteria.

Table 6: Relative weights of sub-criteria with respect to each Criterion

Criteria	Sub-criteria	Relative Weights
Trainer	Idea	0.731
	Participation	0.188
	Sufficient Material	0.081
Content	Subject	0.833
	Handouts	0.167
Time	Timing of Session	0.669
	Training Hours	0.243
	Training Period	0.088
Logistic Facilities	Equipment Availability	0.417
	Comfort	0.417
	Place	0.121
	Hospitality	0.045

These results are expected given the nature of the training programs under study, because most of the training courses offered by these programs are mainly scientific in nature. Therefore, they should be ordered and explained smoothly by trainer. Moreover, most of these courses require running some advanced soft wares. So,

course attendance and participation become important. As for the content criteria, handouts rank very low because they are normally power-point slides, which do not contain many details. On the other hand, for the time criteria, the rank of the timing of the training sessions is high. This can be attributed to the fact that the targeted trainees are either hold jobs or fresh graduates looking for a job.

5.3. Sensitivity Analysis

In order to gain some in-depth insights of the problem, sensitivity analysis is performed in order to study the effect of changing the weights of criteria on the overall weights of the alternatives. The following sections give the details of different types of sensitivity analysis.

5.3.1. Dynamic Sensitivity Analysis Graph

This type of analysis is used to change the weights of the criteria to determine how these changes affect the overall weights of the alternatives. Changing the weights of the criteria depends on the direction in which the criterion is expected to change according to the decision maker. For example, if the decision maker feels that the weight of the trainer criterion might increase from 33% to 50%, while keeping all other criteria fixed, then the weight of alternative a decreases by 4% and that of alternative B increases by 4% as shown in Fig. 4. In this case, alternative B would be the preferred alternative. In general, the decision maker can vary the weight of the criteria so that the training program administration would see its position with respect to competitors.
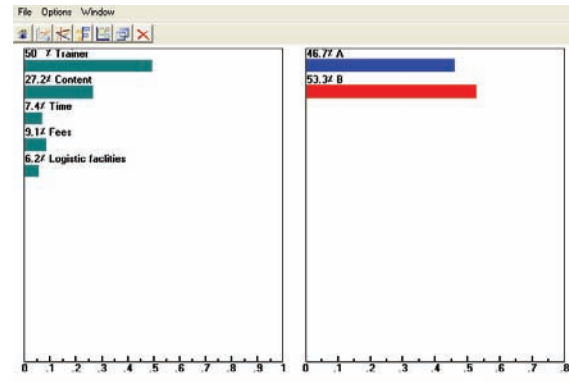

Figure 4: Dynamic sensitivity when the weight of trainer is 50% instead of 33%

5.3.2. Performance Sensitivity Analysis Graph

This type of analysis shows how the alternatives are prioritized relative to other alternatives with respect to each criterion as well as overall. Figure 5. shows the performance sensitivity analysis graph, where X-axis represents the criterion. The height of the bar represents the weight of each criterion. The left Y-axis represents each criterion weight, while the right Y-axis represents the weight of alternatives with respect to each criterion, and the overall weight of each alternative.
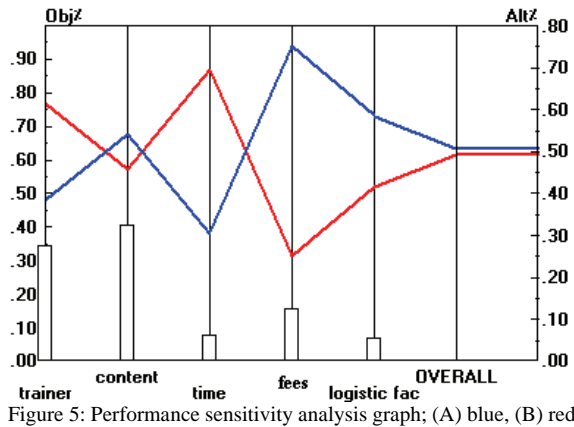
Figure 5: Performance sensitivity analysis graph; (A) blue, (B) red

The advantage of this type is that it represents preference between two alternatives with respect to each criterion. This graph shows the criteria at which the competitor alternative performs better. Therefore, decision maker should focus on the criteria, which are considered weak compared with the competitors. It is clear fro the figure that alternative A outperforms alternative B mainly for the fees and logistic facilities criteria, while, alternative B performs better for the trainer and time criterion. Therefore, alternative A may use the same pool of trainers that is used by alternative B to improve its performance.

5.3.3. Gradient Sensitivity Analysis Graph

This graph shows the alternatives weights with respect to one criterion at a time. The X-axis represents the criterion weight, while the Y-axis represents the overall weights of alternatives. The advantage of this graph is determining whether the decision is sensitive to the change of the criterion weight or not. The decision is sensitive to a given criterion if a small change in the weight of that criterion results in changing in the preferred alternative. Examples of gradient sensitivity analysis are given for some criteria as shown in the next paragraphs.

As for fees criterion, the weight of this criterion may change positively or negatively in the future. If the criterion weight increases, the preferred alternative will not change. Therefore, the decision is not sensitive to the increase in the weight of this criterion. However, if the weight decreases by 4.6%, the preferred alternative will change as shown in Fig. 6. Therefore, the decision is sensitive to the decrease in the weight of fees criterion.
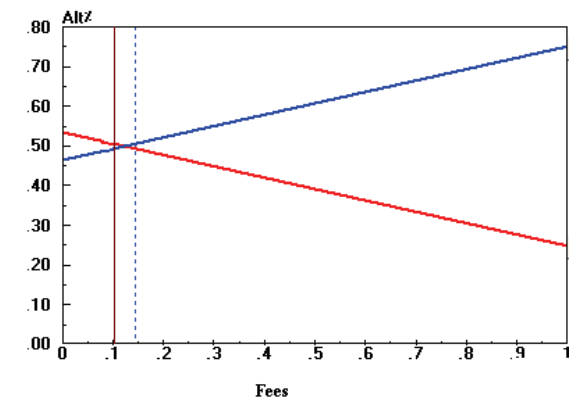

Figure 6: Gradient sensitivity analysis for fees criterion; (A) blue, (B) red

For time criterion, if the criterion weight decreases, the preferred alternative will not change. Therefore, the decision is not sensitive to the decrease in this criterion weight. However, if the criterion weight increases by 8.4%, the preferred alternative will change as shown in Fig. 7. Therefore, the decision is sensitive to the increase in the weight of time criterion. The administration of program B has to watch for changes in the weight of these criteria, otherwise, it may find itself unable to compete with program A and possibly with other existing programs in the field.
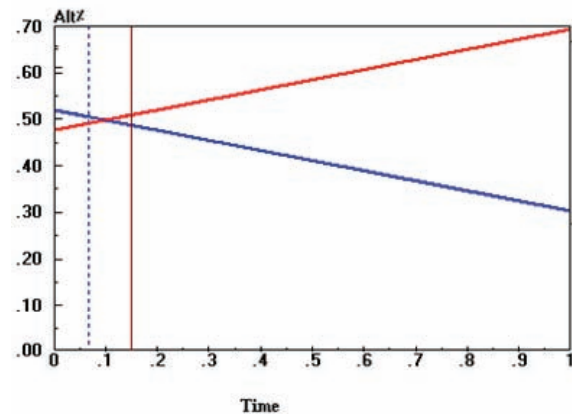

Figure 7: Gradient sensitivity analysis for time criterion; (A) blue, (B) red

For trainer criterion, if the criterion weight decreases, the preferred alternative will not change. Therefore, the decision is not sensitive to the decrease in this criterion weight. However, if the criterion weight increases by 9.1%, the preferred alternative will change as shown in Fig. 8. Therefore, the decision is sensitive to the increase in the weight of trainer criterion.
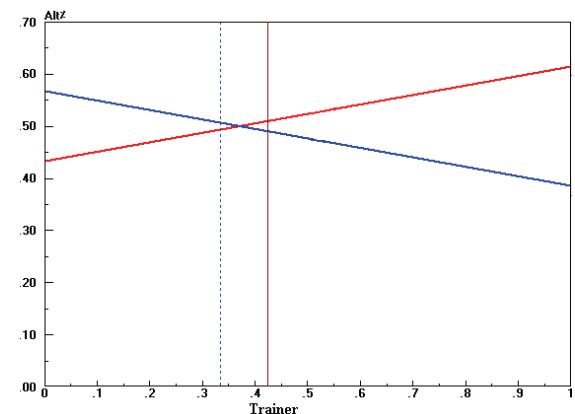

Figure 8: Gradient sensitivity analysis for trainer criterion; (A) blue, (B) red

For content criterion, if the criterion weight increases, the preferred alternative will not change. Therefore, the decision is not sensitive to the increase in this criterion weight. However, if the criterion weight decreases by 2.4%, the preferred alternative will change as shown in Fig. 9. Therefore, the decision is sensitive to the decrease in the weight of content criterion. The fact that such a small change in the weight would reverse the selection of the programs; the administrators of program B have to

watch for small fluctuations, which may easily render their program uncompetitive.
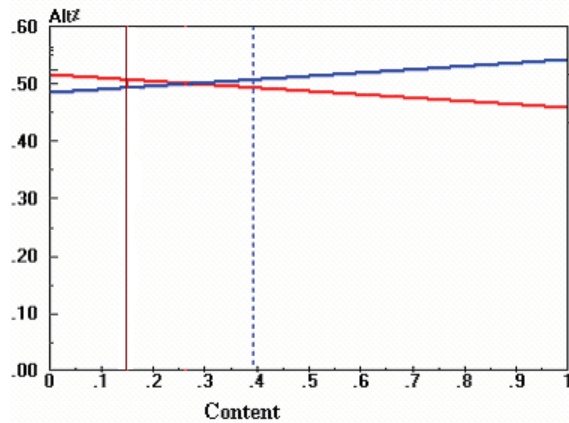

Figure 9: Gradient sensitivity analysis for content criteria; (A) blue, (B) red

6. Results

A hypothetical alternative "C" is derived from the results of alternatives A and B. This alternative "C" can be defined as the alternative that can be obtained if the maximum scores of both alternatives with respect to the sub-criteria are used. In other words, the scores of alternative "C" are the maximum scores obtained for both alternatives A and B. Differently expressed, these maximum values are definitely obtainable in the Gaza Strip environment as evidenced by either alternative A or alternative B. Therefore, either of these program administrators can be assured that the target is obtainable and thus the goals are achievable and consequently taking alternative "C" as an alternative is a more realistic benchmark than just taking one of the alternatives in the given aspect.

Table 10: Relative weight vectors of alternatives with respect to criteria

Alternative	Trainer	Content	Time	Fees	Logistics	Overall Score
A	0.385	0.542	0.305	0.75	0.585	0.506
B	0.615	0.458	0.695	0.25	0.415	0.493
Benchmark (C)	0.75	0.542	0.761	0.75	0.644	0.662

It is clear from Table 10. that alternative A score can be readily increased by 30% if it adopts the best practices of alternatives A and B. While, the score of alternative B can be increased by 34%. The results of the paper would be greatly beneficial for the administrators of these programs. In addition, other similar training programs may benefit from the results by focusing and improving their performance for the criteria, which have high weights. For example, these programs should focus on improving their performances with respect to trainer, time and logistics respectively.

7. Conclusions

1. Critical factors for successful training program are identified. They are trainer, time, fees, content, and logistic facilities. The content and trainer are the most important criteria as they represent 72.8%, where as fees represents 14.5%, while time and logistic facilities are the lowest important criteria as they represent 12.7%.
2. Elements of each critical factor were identified, and ranked with respect to each factor.
3. Strengths and weaknesses are identified for each training program at the case study using sensitivity analysis.
4. A benchmark alternative is derived so that programs can be compared to it.
5. Other similar programs can use the results in order to improve their performances.

Acknowledgement

References

[1]   Almadhoun, M.,"Training under fire: Obstacles facing training and SMEs' development in Palestine". Journal of European Industrial Training, Vol. 30, No. 2, 2006, 100-116.

[2]   Jung, H. W., and Choi, "Optimization models for quality and costs of modular software systems." European Journal of Operational Research, Vol. 112, No. 3, 1999, 613-619.

[3]   Al-harbi, K. M., "Application of AHP in project management." International Journal of Project Management, Vol.19, No.4, 2001, 19-24.

[4]   Zahedi, F, "The analytic hierarchy process: A survey of methods and its applications". Interfaces, Vol. 16, No. 4, 1986, 96-108.

[5]   Cagno, E., Caron, F, and Perego, A., "Multi-criteria assessment of the probability of winning in competitive bidding process." International Journal of Production Management, Vol. 19, 2001, 313-324.

[6]   Tummala, V. R., Chin, K. S., and Ho, S. H., "Assessing success factors for implementing CE: A case study in Hong Kong electronics industry by AHP". International Journal of Production Economics, Vol. 49, No. 3, 1997, 265-283.

[7]   Weiwu, W., and Jun, K., "High way transportation comprehensive evaluation." Computers and Industrial Engineering, Vol. 27, No. 1-4, 1994, 257-259.

[8]   Forgionne, G. A., and Kohli, R., "A multi-criteria assessment of decision technology system and journal quality." Information and Management, Vol. 38, No. 7, 2001, 421-435.

[9]   Badri, M. "Combining AHP and GP for global facility location- allocation problem." International Journal of Production Economics, Vol. 62, No. 3, 1999, 237-248.

[10] Lee, C. W., and Kwak, N. K., "Information resource planning for a health-care system using an AHP-based goal programming method." Journal of Operational Research Society, Vol. 50, 1999, 1191-1198.

[11] Rossetti, M. D., and Selandari, F., "Multi-objective Analysis of hospital delivery systems." Computers and Industrial Engineering, Vol. 41, No. 3, 2001, 309-333.

[12] Koskal, G., and Egitman, A., "Planning, and design of industrial engineering education quality". Computers and Industrial Engineering, Vol. 35, No. 2/3, 1998, 639-642.

[13] Ramanathan, R., and Ganesh, L. S., "Using AHP for resource allocation problems." European Journal of Operational Research, Vol. 80, No. 2, 1995, 410-417.

[14] Shaaban, O., "Reasons of Failure of the Training Programs in Palestine." Al Ayyam Journal, Vol.4, 14-18.

[15] Saaty, T. L., The Analytic Hierarchy Process, 3rd ed. NY: McGraw Hill; 1980,

[16] Expert Choice Inc., "Export Choice, Expert Choice Software, and Manual".. Pittsburgh: 4922 Ellsworth Ave.; 2004