Penilaian Latihan

DCE 3105 (Unit 1-5/5)

Shamsuddin Ahmad, Ph.D
Fakulti Pengajian Pendidikan
Universiti Putra Malaysia
43400 UPM Sedang
Selangor Darul Ehsan
Hak Cipta Terpelihara. Tidak dibenarkan mengeluarkan ulang mana-mana bahagian artikel, ilustrasi dan isi kandungan buku ini dalam apa jua bentuk sama ada secara elektronik, fotokopi,mekanik, rakaman atau cara lain sebelum mendapat izin bertulis daripada Pengarah, Pusat Pendidikan Luar (PPL), Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor Darul Ehsan. Penyertaan serta akan diakibatkan kepada perbuatan royalti atau honorarium.

MODUL PEMBELAJARAN : DCE 3105 PENILAIAN LATHAN disediakan dalam bentuk bahan pengajaran dan pembelajaran krediti di bawah program Pendidikan Jarak Jauh, Universiti Putra Malaysia. Sebarang pertanyaan dan cadangan untuk memperbaiki gaya penyampaian dan isi kandungan modul ini bolehlah dikemukakan kepada penulis dengan menggunakan alamat Pusat Pendidikan Luar.

Penulis : SHAMSUDIN AHMAD, Ph.D
Fakulti Pengajian Pendidikan
Universiti Putra Malaysia
43400 UPM Serdang
Selangor Darul Ehsan

Alamat : Unit Modul dan Bahan Kendiri
Pusat Pendidikan Luar
Universiti Putra Malaysia
43400 UPM Serdang
Selangor Darul Ehsan
Tel : 03-8946830/03-89458904
Fax : 03-8945 8902

UPT DIEN BAHAN KENDIRI DAN MODUL

UPM HOLDINGS SDN. BHD.
Blok F2, Bangunan MTD-UPM
Universiti Putra Malaysia
43400 UPM Serdang
PENILAIAN LATIHAN
(TRAINING EVALUATION)

Oleh
Ahmad Fuad b. Muhammad*.

Pengenalan

Manusia pada umumnya terlibat dalam proses penilaian sama ada secara langsung atau pun tidak dalam kehidupannya setiap hari. Perkara ini dapat dilihat dengan jelas dalam semua urusan yang dilakukan sama ada disedari atau tidak. Contohnya, kita bertanya khabar untuk menilai kesihatan seseorang, kita menilai masa untuk mengetahui waktu, kita bertanya harga barang untuk mengetahui nilainya, dan berbagai perbuatan yang lain seperti meneliti, mengawasi atau mengesahkan. Kesemua ini adalah merupakan penilaian yang seringkali dilakukan dalam bentuk tidak formal. Penilaian juga boleh dilaksanakan dalam bentuk formal isitu dengan cara yang lebih sistematik dan saintifik.

Konsep Penilaian

Penilaian kurang mendapat perenakan sewajarnya dalam proses pembentukan sesuatu program seperti program latihan kakitangan. Salah satu sebab yang seringkali dinostakan ialah kurangnya

* Ketua Unit Perkhidmatan Pengembangan, Pusat Pengembangan dan Pendidikan Lanjutan, Universiti Pertanian Malaysia, Serdang.
pengetahuan tentang konsep-konsep dan kegunaan penilaian yang sebenarnya di kalangan mereka yang terlibat dalam perancangan. Konsep penilaian pada hari ini amatlah terhad skop dan penggunaannya oleh kerana kita sendiri menakrifkan penilaian secara agak sempit dan tidak begitu menyeluruh.

Penilaian secara amnya mengandungi tiga perkara utama iaitu:

i) Kriteria/indikator
   - merupakan 'standard' atau ukuran-ukuran tertentu di mana sesuatu program dinilai.

ii) Bukti (evidence)
    - pengumpulan maklumat mengenai aspek-aspek tertentu.

iii) Membuat pertimbangan (judgement)
    - membandingkan bukti yang dikumpul dengan kriteria yang telah dibentuk.

Perbezaan penilaian formal dengan penilaian tidak formal adalah seperti berikut:

Penilaian Formal
Kriteria --------> Bukti --------> Membuat Pertimbangan

Penilaian Tidak Formal
Membuat pertimbangan --------> Bukti --------> Kriteria
Siapa yang Terlibat dalam Penilaian Latihan?

Siapa yang patut terlibat dalam penilaian latihan bergantung kepada sesuatu aktiviti latihan yang hendak dinilai. Oleh yang demikian sesuatu penilaian yang berjaya dilaksanakan perlu melibatkan beberapa individu atau kumpulan individu. Lazimnya mereka yang terlibat adalah terdiri daripada:

* Pegawai latihan
* Kakitangan yang terlibat dalam latihan
* Pelatih-pelatih (trainee)
* Majikan/penyelia pelatih

Sebagai panduan, individu-individu yang terlibat adalah terdiri daripada mereka yang mempunyai kepentingan untuk mengetahui aktiviti-aktiviti latihan.

Faktor-faktor Utama dalam Penilaian Latihan

Latihan boleh ditaktikkan sebagai suatu perkongsian pengetahuan, sikap dan kemahiran antara jurulatih (trainer) dengan pelatih (trainee). Oleh yang demikian faktor-faktor yang perlu diukur dalam penilaian latihan adalah seperti berikut:

1) Pengetahuan

- bermaksud maklumat-maklumat yang diperolehi oleh pelatih dari jurulatih melalui aktiviti-aktiviti latihan. Pengetahuan pelatih boleh diukur dengan bertanyaan soalan-soalan seperti:
  (a) Berikan definisi/takrif .......?
  (b) Terangkan konsep .......?
  (c) Senaraikan lima faktor .......?
II) Sikap

- bermaksud perasaan yang terbentuk dalam diri pelatih semasa aktiviti latihan. Ia hanya boleh diukur dengan bertanya soalan-soalan seperti;
  
  (a) Bagaimana pendapat anda berkaitan .......?
(b) Sila beri pandangan anda terhadap .......?
(c) Biacangkan konsep .......?

III) Kemahiran

- bermaksud kebolehan/kemampuan yang diperolehi oleh diri pelatih semasa aktiviti latihan. Kemahiran boleh diukur dengan bertanyaan pelatih menunjuk-ajak cara yang betul untuk membiasa sesuatu.

Peringkat-peringkat Penilaian Latihan (rajah 2)

Penilaian latihan boleh dibahagikan kepada beberapa peringkat mengikut faktor-faktor utama (pengetahuan, sikap, kemahiran) yang diperolehi oleh pelatih hasil daripada aktiviti latihan.

1. Penilaian Reaksi (Reaction Evaluation)

Penilaian yang dijalankan di peringkat ini berkehendakkan pelatih-pelatih menyatakan sikapnya terhadap beberapa komponen dalam program latihan seperti;

a) isi kandungan tajuk kursus/berasman,

b) kaedah-latihan dan penggunaan alat bantu mengajar,

c) pentadbiran dan perkhidmatan sokongan seperti penginapan, makanan, pengangkutan, bilik kursus dll.

Hasil daripada penilaian di peringkat ini amat berguna untuk memperbaiki pengurusan latihan dan persekuturnannya dan lazimnya penilaian ini dikendalikan semasa dan syelesa aktiviti latihan.
2. Penilaian Pembelajaran (Learning Evaluation)


3. Penilaian Prestasi (Performance Evaluation)

Penilaian yang dibuat di peringkat ini adalah bertujuan untuk mengukur sejauh mana hasah penggunaan pengetahuan dan kemahiran yang diperoleh oleh pelatih-pelatih di dalam melaksanakan kerja atau tugas harianya. Hasil penilaian sebegini amat berguna untuk mempertingkatkan kemahiran dan produktiviti pelatih. Ianya boleh juga dikendalikan semasa latihan tetapi lazimnya dilaksanakan selepas tamat latihan.

4. Penilaian Keberkesanan (Impact Evaluation)

Penilaian yang dilaksanakan di peringkat ini bertujuan untuk mengukur keasal keseluruhan latihan terhadap prestasi majikan di mana pelatih-pelatih bekerja. Hasil daripada penilaian keberkesanan amat berguna untuk memperbaiki strategi-strategi latihan.
Rajah 2: PERINGKAT DAN PENENKANAN DALAM PENILAIAN LATIHAN

Langkah-langkah Dalam Penilaian Latihan

Penilaian latihan boleh dilaksanakan mengikut langkah-langkah yang sistematis seperti berikut:

* Menentukan jenis penilaian
* Menentukan kaedah penilaian yang hendak digunakan
* Merancang penilaian tersebut
* Menyediakan instrumen penilaian
* Melaporkan hasil penilaian
* Menggunakan hasil penilaian untuk memperbaiki latihan di masa hadapan.
Penutup

Penilaian latihan haruslah diterima sebagai satu amalan yang boleh dilaksanakan dengan sempurna dalam pengurusan latihan. Pada dasarnya, penilaian latihan adalah satu proses pengumpulan maklumat yang bertujuan untuk mempertingkatkan prestasi program pengurusan latihan supaya ianya lebih mantap dan efisien. Berdasarkan konsep ini maka kita tidak keberatan untuk melaksanakan penilaian latihan supaya dapat menghasilkan program-program latihan kakitangan yang berkualiti dan akhirnya untuk mencapai matlamat program pembangunan sumber manusia.

Rujukan


REACTION AND LEARNING EVALUATION QUESTIONNAIRE

Purpose:

1. Reaction Evaluation
   - to know training participants' attitude towards the various components of the training programme

2. Learning Evaluation
   - to know the knowledge, attitude and/or skills gained

Title of Training: ________________________________

Position of Participants: ____________________________

Agency: ____________________________

To help the organizers of the training course improve similar activities, please respond to the following questions. Your answers will be extremely useful in conducting similar training courses in the future.
I. TRAINING PROGRAMME

Please rate (circle) only items which apply.

<table>
<thead>
<tr>
<th>Item</th>
<th>Very Good</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
<th>Very Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Value of training course to your job</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2. Usefulness of the subject matter/content</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3. Effectiveness of training methods used</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4. Usefulness of audio-visual aids used</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5. Speaker(s) ability to transfer knowledge</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6. Atmosphere for active participation</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7. Value of educational visits in the training programme</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>8. Level of instruction</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>9. Lecture coverage</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10. Duration of training course</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11. Clarity of session's objectives</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12. Organisation and direction</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>13. What is the most significant learning experience you have gained from this training?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
II. KNOWLEDGE

(Please provide the correct answers based on discussions during the training).

1. Define the following:
   a) Organisational Management
   
   
   
   b) Managerial Communication
   
   
   
   c) Personnel Management
   
   
   
   d) Evaluation
   
   
   

2. Enumerate some principles of management

   
   
   

3. Enumerate the basic elements of communication

   
   
   


4. Enumerate the four types (levels) of training evaluation

________________________
________________________
________________________

III. ATTITUDE

(Please indicate (✓) wherever appropriate)

<table>
<thead>
<tr>
<th></th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>No Comment</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The main problem in staff development is lack of good management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Decision making is a vital element in organisational management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Communication is important in reducing management conflict</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>The importance of analysing organisational problems is the basis for an effective management system</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IV. SKILL

(Please identify the new skill which you have learned from this training).

1. What new skills did you learn from the training?
2. What other skill do you want to learn in future training?

V. GENERAL

1. On the whole, how would you rate this training course?
   a. Successful
   b. No comment
   c. Failure

2. What are your suggestions to improve similar training course in the future?
PART ONE

CONCEPTS, PRINCIPLES, GUIDELINES, AND TECHNIQUES

Part One contains concepts, principles, guidelines, and techniques for understanding and implementing four levels for evaluating training programs. Most of the content is my own and results from my Ph.D. dissertation on evaluation and my studies and experience since that time. Some modifications were made from the input I received from reviewers that fit in with my objective in writing the book: to provide a simple, practical, four-level approach for evaluating training programs. For those who want information on other principles and techniques, I have provided a selected reading list at the end of Chapter 8.
Chapter 1

Evaluating:
Part of a Ten-Step Process

The reason for evaluating is to determine the effectiveness of a training program. When the evaluation is done, we can hope that the results are positive and gratifying, both for those responsible for the program and for upper-level managers who will make decisions based on their evaluation of the program. Therefore, much thought and planning need to be given to the program itself to make sure that it is effective. Later chapters discuss the reasons for evaluating and supply descriptions, guidelines, and techniques for evaluating at the four levels. This chapter is devoted to suggestions for planning and implementing the program to ensure its effectiveness. More details can be found in my book, How to Train and Develop Supervisors (New York: AMACOM, 1993).

Each of the following factors should be carefully considered when planning and implementing an effective training program:

1. Determining needs
2. Setting objectives
3. Determining subject content
4. Selecting participants
5. Determining the best schedule
6. Selecting appropriate facilities
7. Selecting appropriate instructors
8. Selecting and preparing audiovisual aids
9. Coordinating the program
10. Evaluating the program
Suggestions for implementing each of these factors follow.

Determining Needs

If programs are going to be effective, they must meet the needs of participants. There are many ways to determine these needs. Here are some of the more common:

1. Ask the participants.
2. Ask the bosses of the participants.
3. Ask others who are familiar with the job and how it is being performed, including subordinates, peers, and customers.
4. Test the participants.
5. Analyze performance appraisal forms.

Participants, bosses, and others can be asked in interviews or by means of a survey. Interviews provide more detailed information, but they require much more time. A simple survey form can provide almost as much information and do it in a much more efficient manner.

A survey form, such as the one shown in Exhibit 1.1, can be readily developed to determine the needs seen both by participants and by their bosses. The topics to be considered can be determined by interviews or simply by answering the question, What are all the possible subjects that will help our people to do their best? The resulting list becomes the survey form.

As Exhibit 1.1 indicates, participants are asked to complete the survey by putting a check in one of three columns for each item. This is a much better process than having them list their needs in order of importance or simply writing down the topics that they feel will help them to do their job better. It is important to have them evaluate each topic so that the responses can be quantified.

After you tabulate their responses, the next step is to weight these sums to get a weighted score for each topic. The first column, of great need, should be given a weight of 2; the second column, of some need, should be given a weight of 1; and the last column, a weight of 0. The weighted score can then be used to arrive at a rank order for individual needs. If two topics are tied for third, the next rank is fifth, not fourth, and if three needs have tied for seventh, the next rank is tenth.
In order to determine which subjects will be of the greatest help to you in improving your job performance, we need your input. Please indicate your need for each subject by placing an X in the appropriate column.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Of great need</th>
<th>Of some need</th>
<th>Of no need</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Diversity in the workforce—understanding employees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. How to motivate employees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Interpersonal communications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Written communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Oral communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. How to manage time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. How to delegate effectively</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Planning and organizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Handling complaints and grievances</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. How to manage change</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Decision making and empowerment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Leadership styles—application</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Performance appraisal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Coaching and counseling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. How to conduct productive meetings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Building teamwork</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. How to discipline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Total quality improvement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Safety</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Housekeeping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. How to build morale—quality of work life (QWL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. How to reward performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. How to train employees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. How to reduce absenteeism and tardiness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. Other topics of great need</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This rank order provides training professionals with data on which to determine priorities. Exhibit 1.2 illustrates the tabulations and the rank order.

The same form can be used to determine the needs seen by the bosses of the supervisors. The only change is in the instructions on the form, which should read: “In order to determine which subjects would be of greatest benefit to supervisors to help improve their performance, we need your input. Please put an X in one of the three columns after each subject to indicate the needs of your subordinates as you see them. Tabulations of this survey will be compared with the needs that they see to decide the priority of the subjects to be offered.”

There will be a difference of opinion on some subjects. For example, in a manufacturing organization, the subject of housekeeping might be rated low by supervisors and high by their bosses. Other topics, such as motivation, will probably be given a high rating by both groups. In order to make the final decision on the priority of the subjects to be offered, it is wise to use an advisory committee of managers representing different departments and levels within the organization. The training professional can show the committee members the results of the survey and ask for their input. Their comments and suggestions should be considered advisory, and the training professional should make the final decision.

Participation by an advisory committee accomplishes four purposes:

1. Helps to determine subject content for training programs.
2. Informs committee members of the efforts of the training department to provide practical help.
3. Provides empathy regarding the needs seen by their subordinates.
4. Stimulates support of the programs by involving them in the planning.

The use of tests and inventories is another approach for determining needs. There are two practical ways of doing this. One way is to determine the knowledge, skills, and attitudes that a supervisor should have and develop the subject content accordingly. Then develop a test that measures the knowledge, skills, and attitudes and give it to participants as a pretest. An analysis of the results will provide information regarding subject content.

The other approach is to purchase a standardized instrument that relates closely to the subject matter being taught. The sixty-five-item
### Evaluating: Part of a Ten-Step Process

**Exhibit 1.2. Tabulating Responses to Survey of Training Needs**

In order to determine which subjects will be of the greatest help to you in improving your job performance, we need your input. Please indicate your need for each subject by placing an X in the appropriate column.

<table>
<thead>
<tr>
<th>Rank order</th>
<th>Subject</th>
<th>Weighted score</th>
<th>Of great need</th>
<th>Of some need</th>
<th>Of no need</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1. Diversity in the workforce—understanding employees</td>
<td>40</td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>2. How to motivate employees</td>
<td>51</td>
<td>22</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>3. Interpersonal communications</td>
<td>48</td>
<td>20</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>4. Written communication</td>
<td>33</td>
<td>11</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>23</td>
<td>5. Oral communication</td>
<td>19</td>
<td>6</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>6. How to manage time</td>
<td>44</td>
<td>17</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>7. How to delegate effectively</td>
<td>29</td>
<td>9</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>8. Planning and organizing</td>
<td>29</td>
<td>6</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>14</td>
<td>9. Handling complaints and grievances</td>
<td>39</td>
<td>13</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>10. How to manage change</td>
<td>56</td>
<td>26</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>11. Decision making and empowerment</td>
<td>53</td>
<td>24</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>12. Leadership styles—application</td>
<td>48</td>
<td>19</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>13. Performance appraisal</td>
<td>36</td>
<td>12</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>14. Coaching and counseling</td>
<td>36</td>
<td>8</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>15. How to conduct productive meetings</td>
<td>27</td>
<td>8</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>16. Building teamwork</td>
<td>55</td>
<td>25</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>17. How to discipline</td>
<td>47</td>
<td>18</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>18. Total quality improvement</td>
<td>39</td>
<td>13</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>19. Safety</td>
<td>43</td>
<td>15</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>20. Housekeeping</td>
<td>19</td>
<td>6</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>21. How to build morale—quality of work life (QWL)</td>
<td>50</td>
<td>22</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>22. How to reward performance</td>
<td>41</td>
<td>17</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>23. How to train employees</td>
<td>48</td>
<td>19</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>24. How to reduce absenteeism and tardiness</td>
<td>31</td>
<td>11</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>25</td>
<td>Other topics of great need</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Tabulated responses from thirty first-level supervisors.*
Evaluating Training Programs

Management Inventory on Managing Change (available from Donald L. Kirkpatrick, 1920 Hawthorne Drive, Elm Grove, WI 53122) is such an instrument. Here are some of the items in it:

1. If subordinates participate in the decision to make a change, they are usually more enthusiastic in carrying it out.
2. Some people are not anxious to be promoted to a job that has more responsibility.
3. Decisions to change should be based on opinions as well as on fact.
4. If a change is going to be unpopular with your subordinates, you should proceed slowly in order to obtain acceptance.
5. It is usually better to communicate with a group concerning a change than to talk to its members individually.
6. Empathy is one of the most important concepts in managing change.
7. It's a good idea to sell a change to the natural leader before trying to sell it to the others.
8. If you are promoted to a management job, you should make the job different than it was under your predecessor.
9. Bosses and subordinates should have an understanding regarding the kinds of changes that the subordinate can implement without getting prior approval from the boss.
10. You should encourage your subordinates to try out any changes that they feel should be made.

Respondents are asked to agree or disagree with each statement. The "correct" answers were determined by the author to cover concepts, principles, and techniques for managing change. It is important to note that the possible answers are "agree" or "disagree" and not "true" or "false."

Five other standardized inventories are available from the source just named: Supervisory Inventory on Communication, Supervisory Inventory on Human Relations, Management Inventory on Time Management, Management Inventory on Performance Appraisal and Coaching, and Management Inventory on Leadership, Motivation, and Decision Making.

Many other approaches are available for determining needs. Two of the most practical—surveying participants and their bosses and
giving a pretest to participants before the program is run—have just been described.

Setting Objectives

Once the needs have been determined, it is necessary to set objectives. Objectives should be set for three different aspects of the program and in the following order:

1. What results are we trying to accomplish? These results can be stated in such areas as production, sales, quality, turnover, absenteeism, and morale or quality of work life (QWL). Some organizations set objectives in terms of profits, even on return on investment (ROI). This is usually a mistake, because so many factors determine results like these that evaluation in these terms is often impossible.

2. What behaviors do we want supervisors and managers to have in order to accomplish the results? One such behavior may be management by walking around (MBWA), the concept described by Thomas Peters and Robert Waterman, Jr., in In Search of Excellence (New York: Warner Books, 1982). All levels of management at United Airlines and Hewlett-Packard are urged to use this approach to show employees that they care about them. The desired results are better quality of work life, higher morale, and thereby improved productivity.

3. What knowledge, skills, and attitudes do we want participants to learn in the training program? Some programs are aimed at teaching specific knowledge or skills. Others, such as programs on diversity in the workforce, are aimed at increasing knowledge and changing attitudes.

Robert Mager's book Preparing Instructional Objectives (Belmont, Calif.: Lake, 1962) describes specific concepts and approaches. (See references at end of Chapter 8.)

Determining Subject Content

Needs and objectives are prime factors when determining subject content. Trainers should ask themselves the question, What topics should be presented to meet the needs and accomplish the objectives? The answers to this question establish the topics to be covered. Some modifications may be necessary depending on the qualifications of the
trainers who will present the program and on the training budget. For example, the subject of managing stress may be important, but the instructors available are not qualified, and there is no money to hire a qualified leader or buy videotapes and/or packaged programs on the subject. Other pertinent topics then become higher priorities.

Selecting Participants

When selecting participants for a program, four decisions need to be made:

1. Who can benefit from the training?
2. What programs are required by law or by government edict?
3. Should the training be voluntary or compulsory?
4. Should the participants be segregated by level in the organization, or should two or more levels be included in the same class?

In answer to the first question, all levels of management can benefit from training programs. Obviously, some levels can benefit more than others. The answer to the second question is obvious. Regarding the third question, I recommend that at least some basic programs be compulsory for first-level supervisors if not also for others. If a program is voluntary, many who need the training may not sign up, either because they feel they don't need it or because they don't want to admit that they need it. Those who are already good supervisors and have little need for the program can still benefit from it, and they can also help to train the others. This assumes, of course, that the program includes participatory activities on the part of attendees. To supplement the compulsory programs, other courses can be offered on a voluntary basis.

Some organizations have established a management institute that offers all courses on a voluntary basis. Training professionals may feel that this is the best approach. Or higher-level management may discourage compulsory programs. If possible, the needs of the supervisors, as determined by the procedures described in the preceding section, should become basic courses that should be compulsory. Others can be optional. The answer to the last question depends on the climate and on the rapport that exists among different levels of management within the organization. The basic question is whether sub-
ordinates will speak freely in a training class if their bosses are present. If the answer is yes, then it is a good idea to have different levels in the same program. They all get the same training at the same time. But if the answer is no, then bosses should not be included in the program for supervisors. Perhaps you can give the same or a similar program to upper-level managers before offering it to the first-level supervisors.

**Determining the Best Schedule**

The best schedule takes three things into consideration: the trainees, their bosses, and the best conditions for learning. Many times, training professionals consider only their own preferences and schedules. An important scheduling decision is whether to offer the program on a concentrated basis—for example, as a solid week of training—or to spread it out over weeks or months. My own preference is to spread it out as an ongoing program. One good schedule is to offer a three-hour session once a month. Three hours leave you time for participation as well as for the use of videotapes and other aids. The schedule should be set and communicated well in advance. The day of the program and the specific time should be established to meet the needs and desires of both the trainees and their bosses. Line managers should be consulted regarding the best time and schedule.

I recently conducted a week-long training program for all levels of management at a company in Racine, Wisconsin. Two groups of twenty each attended the program. The first session each day was scheduled from 7:00 to 10:30 A.M. The repeat session for the other group was scheduled from 3:00 to 6:30 P.M. Racine was too far away to go home each day, and what do you do in Racine from 10:30 A.M. to 3:00 P.M. each day for a week? This is the worst schedule I ever had, but it was the best schedule for all three shifts of supervisors who attended. The point is, the training schedule must meet the needs and desires of the participants instead of the convenience of the instructors.

**Selecting Appropriate Facilities**

The selection of facilities is another important decision. Facilities should be both comfortable and convenient. Negative factors to be avoided
include rooms that are too small, uncomfortable furniture, noise or other distractions, inconvenience, long distances to the training room, and uncomfortable temperature, either too hot or too cold. A related consideration has to do with refreshments and breaks. I conducted a training program on managing change for a large Minneapolis company. They provided participants with coffee and sweet rolls in the morning, a nice lunch at noon, and a Coke and cookie break in the afternoon. Participants came from all over the country, including Seattle. In order to save money on transportation and hotel, the company decided to take the program to Seattle, where it had a large operation. In Seattle, no refreshments were offered, and participants were on their own for lunch. Unfortunately, some peers of the participants had attended the same program in Minneapolis. These factors caused negative attitudes on the part of those attending. And these attitudes could have affected their motivation to learn as well as their feeling toward the organization and the training department in particular. Incidentally, more and more companies are offering fruit instead of sweet rolls and cookies at breaks.

Selecting Appropriate Instructors

The selection of instructors is critical to the success of a program. Their qualifications should include a knowledge of the subject being taught, a desire to teach, the ability to communicate, and skill at getting people to participate. They should also be "learner oriented"—have a strong desire to meet learner needs.

Budgets may limit the possibilities. For example, some organizations limit the selection to present employees, including the training director, the Human Resources manager, and line and staff managers. There is no money to hire outside leaders. Therefore, subject content needs to be tailored to the available instructors, or else instructors need to receive special training. If budgets allow, outside instructors can be hired if internal expertise is not available. The selection of these instructors also requires care. Many organizations feel that they have been burned because they selected outside instructors who did a poor job. In order to be sure that a potential instructor will be effective, the best approach is to observe his or her performance in a similar situation. The next best approach is to rely on the recommendations of
other training professionals who have already used the individual. A very unreliable method is to interview the person and make a decision based on your impressions.

I recently conducted a workshop for eighty supervisors and managers at St. Vincent Hospital in Indianapolis. I had been recommended to Frank Magliery, vice president of Operations, by Dave Neil of ServiceMaster. Dave had been in several of my sessions. In order to be sure that I was the right instructor, Frank attended another session that I did for ServiceMaster. He was able therefore not only to judge my effectiveness but also to offer suggestions about tailoring the training to his organization.

This is the kind of selection process that should be followed when you hire an outside consultant. It not only illustrates a process for selection but also emphasizes the importance of orienting an outside leader to the needs and desires of the specific organization.

Selecting and Preparing Audiovisual Aids

An audiovisual aid has two purposes: to help the leader maintain interest and to communicate. Some aids, hopefully only a few minutes long, are designed to attract interest and entertain. This is fine providing they develop a positive climate for learning. When renting or purchasing videotapes and packaged programs, take care to preview them first to be sure that the benefits for the program outweigh the cost. The extent to which such aids should become the main feature of a program depends on the instructor’s knowledge and skills in developing his or her own subject content. Some organizations rely entirely on packaged programs because they have the budget but not the skills needed to develop and teach programs of their own. Other training professionals rely primarily on their own knowledge, skill, and materials and rent or buy videos only as aids. Some organizations have a department that can make effective aids and provide the necessary equipment. Other organizations have to rent or buy them. The important principle is that aids can be an important part of an effective program. Each organization should carefully make or buy the aids that will help it to maintain interest and communicate the message.
Coordinating the Program

Sometimes the instructor coordinates as well as teaches. In other situations a coordinator does not do the teaching. For those who coordinate and do not teach, there are two opposite approaches.

As an instructor, I have experienced two extremes in regard to coordination. At an eastern university offering continuing education, I had to introduce myself, find my way to the lunchroom at noon, tell participants where to go for breaks, conclude the program, and even ask participants to complete the reaction sheets. I couldn’t believe that a university that prided itself on professional programming could do such a miserable job of coordinating.

The other extreme occurred in a program that I conducted for State Farm Insurance in Bloomington, Illinois. Steve Whittington and his wife took my wife, Fern, and me out to dinner the evening before the program. He picked me up at the hotel to take me to the training room in plenty of time to set the room up for the meeting. He made sure that I had everything I needed. He introduced me and stayed for the entire program, helping with handouts. He handled the breaks. He took me to lunch and, of course, paid for it. He concluded the meeting by thanking me and asking participants to complete reaction sheets. He took me back to the hotel and thanked me. In other words, he served as an effective coordinator who helped to make the meeting as effective as possible. Of course, the niceties that he included are not necessary for effective coordination, but they do illustrate that it is important to meet the needs of the instructor as well as of the participants.

Evaluating the Program

Details on evaluation are provided in the rest of the book.

As stated at the beginning of this chapter, to ensure the effectiveness of a training program, time and emphasis should be put on the planning and implementation of the program. These are critical if we are to be sure that, when the evaluation is done, the results are positive. Consideration of the concepts, principles, and techniques described in this chapter can help to ensure an effective program.
Chapter 2

Reasons for Evaluating

At a national conference of the National Society for Sales Training Executives (NSSTE), J. P. Huller of Hobart Corporation presented a paper on "evaluation." In the introduction, he says, "All managers, not just those of us in training, are concerned with their own and their department's credibility. I want to be accepted by my company. I want to be trusted by my company. I want to be respected by my company. I want my company and my fellow managers to say, 'We need you.'

"When you are accepted, trusted, respected, and needed, lots and lots of wonderful things happen:

- Your budget requests are granted.
- You keep your job. (You might even be promoted.)
- Your staff keep their jobs.
- The quality of your work improves.
- Senior management listens to your advice.
- You're given more control.

"You sleep better, worry less, enjoy life more. . . . In short, it makes you happy.

"Wonderful! But just how do we become accepted, trusted, respected, and needed? We do so by proving that we deserve to be accepted, trusted, respected, and needed. We do so by evaluating and reporting upon the worth of our training."

15
Evaluating Training Programs

This states in general terms why we need to evaluate training. Here are three specific reasons:

1. To justify the existence of the training department by showing how it contributes to the organization’s objectives and goals
2. To decide whether to continue or discontinue training programs
3. To gain information on how to improve future training programs

There is an old saying among training directors: When there are cutbacks in an organization, training people are the first to go. Of course, this isn’t always true. However, whenever downsizing occurs, top management looks for people and departments that can be eliminated with the fewest negative results. Early in their decision, they look at such overhead departments as Human Resources. Human Resources typically includes people responsible for employment, salary administration, benefits, labor relations (if there is a union), and training. In some organizations, top management feels that all these functions except training are necessary. From this perspective, training is optional, and its value to the organization depends on top executives’ view of its effectiveness. Huller is right when he states that training people must earn trust and respect if training is to be an important function that an organization will want to retain even in a downsizing situation. In other words, trainers must justify their existence. If they don’t and downsizing occurs, they may be terminated, and the training function will be relegated to the Human Resources manager, who already has many other hats to wear.

The second reason for evaluating is to determine whether you should continue to offer a program. The content of some programs may become obsolete. For example, programs on Work Simplification, Transactional Analysis, and Management by Objectives were “hot” topics in past years. Most organizations have decided to replace these with programs on current hot topics such as Diversity, Empowerment, and Team Building. Also, some programs, such as computer training, are constantly subject to change. Some programs are offered on a pilot basis in hopes that they will bring about the results desired. These programs should be evaluated to determine whether they should be continued. If the cost outweighs the benefits, the program should be discontinued or modified.

The most common reason for evaluation is to determine the effectiveness of a program and ways in which it can be improved. Usually,
the decision to continue it has already been made. The question then is, How can it be improved? In looking for the answer to this question, you should consider these eight factors:

1. To what extent does the subject content meet the needs of those attending?
2. Is the leader the one best qualified to teach?
3. Does the leader use the most effective methods for maintaining interest and teaching the desired attitudes, knowledge, and skills?
4. Are the facilities satisfactory?
5. Is the schedule appropriate for the participants?
6. Are the aids effective in improving communication and maintaining interest?
7. Was the coordination of the program satisfactory?
8. What else can be done to improve the program?

A careful analysis of the answers to these questions can identify ways and means of improving future offerings of the program.

When I talked to Matt, a training director of a large bank, and asked him to write a case history on what his organization has done to evaluate its programs, here is what he said: "We haven’t really done anything except the ‘smile’ sheets. We have been thinking a lot about it, and we are anxious to do something. I will be the first one to read your book!"

This is the situation in many companies. They use reaction sheets (or "smile" sheets, as Matt called them) of one kind or another. Most are thinking about doing more. They haven’t gone any further for one or more of the following reasons:

- They don’t consider it important or urgent.
- They don’t know what to do or how to do it.
- There is no pressure from higher management to do more.
- They feel secure in their job and see no need to do more.
- They have too many other things that are more important or that they prefer to do.

In most organizations, both large and small, there is little pressure from top management to prove that the benefits of training outweigh the cost. Many managers at high levels are too busy worrying about profits, return on investment, stock prices, and other matters of concern
to the board of directors, stockholders, and customers. They pay little or no attention to training unless they hear bad things about it. As long as trainees are happy and do not complain, trainers feel comfortable, relaxed, and secure.

However, if trainees react negatively to programs, trainers begin to worry, because the word might get to higher-level managers that the program is a waste of time or even worse. And higher-level managers might make decisions based on this information.

In a few organizations, upper-level managers are putting pressure on trainers to justify their existence by proving their worth. Some have even demanded to see tangible results as measured by improvements in sales, productivity, quality, morale, turnover, safety records, and profits. In these situations, training professionals need to have guidelines for evaluating programs at all four levels. And they need to use more than reaction sheets at the end of their programs.

What about trainers who do not feel pressure from above to justify their existence? I suggest that they operate as if there were going to be pressure and be ready for it. Even if the pressure for results never comes, trainers will benefit by becoming accepted, respected, and self-satisfied.

Summary

There are three reasons for evaluating training programs. The most common reason is that evaluation can tell us how to improve future programs. The second reason is to determine whether a program should be continued or dropped. The third reason is to justify the existence of the training department. By demonstrating to top management that training has tangible, positive results, trainers will find that their job is more secure, even if and when downsizing occurs. If top-level managers need to cut back, their impression of the need for a training department will determine whether they say, “That’s one department we need to keep” or “That’s a department that we can eliminate without hurting us.” And their impression can be greatly influenced by trainers who evaluate at all levels and communicate the results to them.
PROGRAM EVALUATION: THEORY AND PRACTICE

BY

Dr. Shamsuddin Ahmad
Department of Professional Development and Continuing Education
Faculty of Educational Studies
Universiti Putra Malaysia
September 2002

Introduction

Theory and practice are the two domains of the program evaluation field. Both are intended to be synonymous (Forest, 1976). However, in reality there is a gap between theory and practice of program evaluation. On one hand, theory suggests the use of a formal, systematic, and comprehensive approach to evaluation, whereas in practice, studies have shown that practitioners do not follow the theory. There is no pure type of evaluation following one specific theory (Chen and Rossi, 1992). There are many factors suggested for the disjuncture; however, no empirical study examining the factors in detail has been found. This paper will explore, integrate, and summarize program evaluation literature in both the general evaluation field and the adult and continuing education field, particularly literature pertinent to theory and practice which are related with the interest of this paper.

This review of the literature will help to conceptualize the theory and practice of program evaluation. It is divided into three topics which are organized according to a logical flow of the themes. The topics are: (1) evaluation theory and model; (2) evaluation in adult education; (3) the practice of program evaluation; (4) the practice of evaluation in adult education; (5) suggested factors influencing evaluation practices; and (6) evaluation and adult education in Malaysia.
Evaluation Theory and Models

There is an important symbiotic relationship between the theory and the practice of program evaluation. Theory is developed to guide practice and practice should be used to refine theory (Chen and Rossi, 1992). Evaluation’s theoretical framework helps practitioners define evaluation problems and design strategies for resolving the problems. "Theory connotes a body of knowledge that organizes, categorizes, describes, predicts, explains, and otherwise aids in understanding and controlling a topic” (Shadish, Cook, and Leviton, 1991, p. 30). Whereas, a model is an extension of theory which includes the "how to" components. According to Madaus, Scriven, and Stufflebeam (1983), an evaluation model presents "the main concepts involved in evaluation work and provides guidelines for using these concepts to arrive at defensible descriptions, judgments, and recommendations” (p. xii-xiii). In addition, the authors describe models as idealized views on how to sort and address problems encountered in conducting evaluations. In general there are more then fifty program evaluation models that have gained some acceptance found in the literature (Steele, 1989).

According to Shadish, Cook, and Leviton (1991), a good comprehensive evaluation theory must explicitly describe and justify at least five fundamental components which are (p. 32):

1. Social programming: the ways that social programs and policies develop, improve, and change, especially in regard to social problems

2. Knowledge construction: the ways researchers learn about social action

3. Valuing: the ways value can be attached to program descriptions

4. Knowledge use: the ways social science information is used to modify programs and policies
5. Evaluation practice: the tactics and strategies evaluators follow in their professional work, especially given the constraints they face.

The authors further divided evaluation theory into three stages. The first stage deals with concepts for valuing and knowledge construction which advocate a rigorous scientific methodology. The next stage is more concerned with being more realistic about the nature of social programs and how it applies in policy development. And finally, the third stage tries to integrate the alternatives generated from the previous two stages.

In an attempt to systematically describe and analyze major evaluation models, House (1978) classifies the models into eight general categories based on their theoretical assumptions. According to the author, the assumptions simply mean "things taken for granted" or "things taken to be true" (p. 45). At one level of abstraction, the assumptions consist of the evaluation's audiences, what the model assumes consensus on, data collection's methodology, the expected outcomes, and what questions the models try to address. The models are system analysis, behavioral objectives, decision-making, goal-free, art criticism, accreditation, adversary, and transaction. The brief assumptions of the models are described below.

In the system analysis approach, the model "assumes a few quantitative output measures, usually test scores, and tries to relate differences in programs to variations in test scores" (p. 46). Usually this model uses survey and experimental methodology. In the behavioral objectives model the "objectives of a program are spelled out in terms of specific student performances that can be reduced to specific student behaviors" (p. 46). This model uses test scores to measure the behaviors, using norm references or criterion references as standards. The decision-making model
is "structured by the decisions to be made" (p. 46). Evaluators supply the information to administrators and managers for making particular decisions. This model usually employs questionnaire and interview surveys. On the other hand, goal-free evaluation model is "concerned with reducing the effects of bias in evaluation" (p. 46) by searching all outcomes. The primary audience for this model are the consumers. The art criticism model or educational critic uses the traditions of art and literary criticism. This model uses an expert "who is attuned by experience and training to judge the important facets of educational programs" (p. 46). The task of the critic is to help us to judge the programs through the process of description, interpretation, and appraisal (Eisner, 1983). Similarly, the accreditation evaluation model employs outside professionals to judge programs using a set of external standards. This is a cooperative effort by schools to review their programs. On the other hand, the adversary evaluation model uses quasi-legal procedures by presenting pros and cons of a program. This method usually takes the form of trial-by-jury to determine the worth of a program. Last, but not least, is the transaction model. Evaluation using this model puts emphasis on the process rather than on the outcome of a program. In addition, this model also "focuses on events occurring in and around the actual program in context. It is based on perception and knowing as a transactional process" (p. 57-58). A typical question for this model is what does the program look like to different people? Frequently, the model uses case study, interview, and observation methodologies.

Another analysis of evaluation models is carried out by Guba and Lincoln (1989). The authors divide evaluation models into four generations characterized by different methodologies. They are: measurement-oriented; description-oriented; judgment-oriented; and negotiation-oriented, a model developed by the authors.

Guba and Lincoln (1989) use the term generation to indicate that each new generation of a
model was build upon previously existing generations. First generation evaluation, measurement-oriented, was developed to determine whether students have "mastered" the contents of educational programs. Students' performances are measured using test scores. The model is proliferated by the development of the scientific approach to studying social phenomena. During this era, the term evaluation and measurement are used interchangeably. The second generation evaluation, description-oriented, was developed due to the limited use of the measurement-oriented model which only targeted the student as a source of data. Such data is not enough in order to improve existing programs. This second generation model measures the achievement of desired learning outcomes, also known as objectives. In addition this model describes strengths and weakness with respect to certain stated objectives. Evaluators in this model take the role of describer while still holding the technical role which characterized the previous first generation model. Tyler's objective-oriented model is one example of the second generation evaluation. The third generation evaluation, judgment-oriented, was built in order to improve the second generation model. This third generation model includes efforts to reach judgments about the worth of programs using certain criteria as standards. Some of the models included in this generation are Stake's countenance model, Provus' Discrepancy evaluation model, Stufflebeam's CIPP model, Scriven's goal-free model, and Eisner's art criticism model.

The fourth generation model, negotiation-oriented, which was developed by Guba and Lincoln (1989) is based on a naturalistic paradigm. The model emphasizes the need to negotiate with stakeholders regarding their claims, concerns, and issues. This responsive mode of focusing is the source of information to be used in the evaluation process. The authors suggest the use of constructivist methodology consisting of twelve steps vicious cycle of working with stakeholders. The cycle starts with initiating a contract with the client/sponsor and end with making an
evaluation report to client/sponsor. Even though the fourth generation model does not follow the positivist paradigm, the evaluation design suggested by the authors is in line with a formal, systematic, and comprehensive approach of evaluation.

House (1978), and Guba and Lincoln (1989) all agree that the existing evaluation models are connected in some way. House sees that a systematic linkage exists between the eight models he analyzed. For example, the methodology of the models moves from a more quantitative to a more qualitative paradigm when one moves from the system-analysis model to the transaction model. Likewise, Guba and Lincoln (1989) also see the relationship between the models. "Each succeeding generation represented a step forward, both in the range of substance or content included in the construction held as well as in its level of sophistication" (p. 31).

Likewise, Stufflebeam and Shinkfield (1985) classified evaluation models into three broad types. There are pseudoevaluation, quasievaluation, and true evaluation. Pseudoevaluation "includes attempts to mislead through evaluation" (p. 45). The objective of this type of evaluation approach is to create a false impression of an object's value. Some examples of this approach include covert investigations and public-relations-inspired studies. According to Stufflebeam and Shinkfield (1985), pseudoevaluation unfortunately is a dominant part of the evaluation scene. Besides the deception, this model is not comprehensive. The authors do not recommend these studies. On the other hand, quasievaluation "encompasses studies that are preoccupied with answering given questions of interest instead of assessing the value of something" (p. 45). Even though quasievaluation sometimes happens to provide enough evidence to assess the worth or merit of a program, it is usually too narrow in focus. Some examples of this model are objective-based studies, and experiment-oriented studies.

True evaluation employs formal and systematic methods of data collection and analysis.
In addition, the authors only consider true evaluation as "comprehensive efforts to examine the worth and merit of an object" (p.45). True evaluation emphasizes the need to guard against biased studies. Some examples of a true evaluation model suggested by Stufflebeam and Shinkfield are Cronbach's designing evaluation, Stufflebeam's improvement-oriented evaluation, Stake's client-centered evaluation, Owen and Wolf's adversary evaluation, Scriven's consumer-oriented evaluation, MacDonald and Stake's holistic approach, and Weiss's politics and interpersonal approaches. For example, Cronbach's designing evaluation model, among other things, is designed to provide clear, timely, reliable, valid, and wide-ranging information that fits in the assertment of a formal, systematic, and comprehensive evaluation approach.

Using Stufflebeam and Shinkfield's (1985) schemes of classification, evaluation models analyzed by House (1978) and Guba and Lincoln (1989) fit in either quasievaluation or true evaluation categories. In terms of methodology, Stufflebeam and Shinkfield (1985) prefer the use of an eclectic approach. Evaluators should be proficient with a wide range of techniques of evaluation design and use the most suitable design to suit the evaluation context. In sum, Stufflebeam and Shinkfield (1985) and Guba and Lincoln (1989) advocate the use of formal, systematic, and comprehensive approaches of program evaluation.

In an effort to maintain high quality evaluation work, certain standards have been suggested. The Evaluation Research Society, The Joint Committee on Standards for Educational Evaluation, and the General Accounting Office have developed some standards for evaluation practices (Rossi, 1982). The standards are in line with the formal, systematic, and comprehensive approach of program evaluation suggested by the literature. For example, The Joint Committee on Standards for Educational Evaluation (1981) developed four categories of standards which are utility standards, feasibility standards, propriety standards, and accuracy standards. According to
the committee, utility standards are intended "to ensure that an evaluation will serve the practical information needs of given audiences" (p. 19). The standards include audience identification, evaluator credibility, information scope and selection, valuational interpretation, report clarity, report dissemination, report timeliness, and evaluation impact. The second category, feasibility standards, are designed "to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal" (p. 51). Considerations involved are practical procedures, political viability, and cost effectiveness. The third category, propriety standards, are intended "to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results" (p. 63). These standards are formal obligation, conflict of interest, full and frank disclosure, the public's right to know, rights of human subjects, human interactions, balanced reporting, and fiscal responsibility. Lastly, the accuracy standards category is designed "to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its worth or merit" (p. 99). These standards are object identification, context analysis, described purposes and procedures, defensible information sources, valid measurement, reliable measurement, systematic data control, analysis of quantitative information, analysis of qualitative information, justified conclusions, and objective reporting.

Theory and model are developed as guides to evaluation practices. Even though there are no pure type of evaluation practices following one specific theory or model (Chen and Rossi, 1992), the practices do have some elements rooted in one of the theories or models (Fetterman, 1992). Practitioners prefer literature on evaluation models to evaluation theory because evaluation models have the 'how to' components (Cronbach and assoc., 1980). There are many different evaluation models that have gained some acceptance found in the literature. In general,
most of the models assert the use of formal, systematic, and comprehensive approaches of program evaluation. In addition, evaluation standards are also being proposed which are in line with the above approach.

Evaluation in Adult Education

"There does not appear to be much development work taking place concerning evaluation in adult education" (Stakes, 1981, p. 23). For example, in the analysis of program planning literature (Sork and Busky, 1986), and the analysis of administration literature (Courtenay, 1990) program evaluation is not extensively treated in the field of adult education. As a result, adult educators have to look at other educational and social program evaluations literature. However, it is not sufficient to follow evaluation traditions from other fields and care should be taken in choosing which model is suitable to adult education's philosophy and beliefs (Steele, 1989).

Using Stufflebeam's framework of conceptualization of evaluation, Stakes (1981) suggests these following statements to characterize evaluation in adult education: goal attainment is the dominant aspect of defining evaluation; the purpose of evaluation is to determine the effectiveness of intervention in achieving its objectives; the emphasis is on summative evaluation; evaluation is viewed as administrative tools to assess program effectiveness; and there is no effective criteria in metaevaluation in adult education (p.23). The author finds that adult education evaluation literature realizes the importance of dealing with major areas of program development and implementation in order to determine the extent to which objectives have been achieved.

Steele (1989) favors a comprehensive approach of program evaluation in the adult education field. The author suggested usable evaluations for evaluating adult and continuing
education programs. The evaluation model suggested by the author is the combination of Chen and Rossi's theory-driven, Stufflebeam's CIPP (Context, Input, Process, and Product), and Wholey's evaluability assessment. Evaluability assessment is a preevaluation process using a systematic approach of providing a climate favorable for future evaluation activities. The process examines the program to ensure it can be evaluated adequately by determine (1) whether the program has been carried out as planned and what changes must be taken into consideration in the evaluation, (2) whether stakeholders agree on the purpose of the program and of the evaluation, (3) whether the program design is likely to yield the results expected, and (4) whether there is likely to be sufficient use of evaluation findings to justify the cost (Steele, 1989, p. 265).

The suggested models require adult educators to have special skills in mental and social process, knowledge in categories of educational change, and knowledge in program theory. The author suggests that in order to create a greater understanding of evaluation results, all important aspects of the program development process should be assessed.

In addition, Steele (1989) suggests that evaluation in adult education needs more than attainment of objectives as the purpose of evaluation. Among other things, it needs to look at both expected and unexpected results as well as negative and positive results. The focus of evaluation also needs to be holistic. It requires adult educators to focus on proof of effect, judgment against criteria, critical questions, and valuing. In terms of the methodology, the author suggests the use of multiple methods. "Evaluation needs to encompass diverse activities and to be designed to fit specific users and evaluation situations" (p.267). Finally, the author proposes that adult education evaluation requires much broader standards. Besides the scientific standard, it requires utility standards such as timeliness - report must be timely for best used by the audience, relevance - report must be pertinent to the need of the audience, understandability - report must
be clear in terms of context, purpose, procedure, and finding, credibility - evaluators must be trustworthy and competent, and usefulness - evaluation must be designed to provide findings that fulfill stakeholder's needs.

Grotelueschen (1980) also provides a comprehensive approach of program evaluation. The author suggests that evaluation not only assess the purpose but also the trends and preferences of the audiences, the nature of the program, and the context where it operates. Grotelueschen asserts that evaluation should consider the totality of program variables. The program variables include (p. 77): general program descriptors (e.g., type of agency, staff characteristics, etc.), explicitly stated program goals, program states (how program is viewed at different time), and central concerns for and about the program (procedural concerns expressed during implementation).

In continuing education, Deshler (1984) suggests more systematic evaluations be used for a variety of alternative purposes such as accountability purposes, practice-improvement purposes, and social learning purposes. The author views models as carpenter's tool. "It is ridiculous for carpenters to argue over whether saws are better than hammers" (p. 12). The tools are designed for different purposes. It is up to the evaluators to properly choose which model suits their needs, purposes, and context.

The effectiveness of a continuing education program is not widely measured in terms of impact evaluation. Assessment is mainly concerned with what happens at a program. In order to do an effective evaluation, it is important to assess knowledge utilization following instruction and training (Holt and Courtenay, 1985). As a result, Courtenay and Holt (1987) argue for a more formal and systematic approach to program evaluation as an alternative for the informal method called "satisfaction indexes" or "happiness quotient". The authors suggest the use of impact
evaluation which should be appropriately planned and built into the program development process. This is because some aspects of program development are important in order to determine the impact of continuing education programs.

Likewise, Cervero (1988) also proposes a more systematic and comprehensive evaluation procedure in the field of continuing professional education (CPE). Evaluation should not be separated from the program development process. "To be effective, evaluation process must be embedded in the cycle of program development" (p. 133). Every important aspect of program development ought to be evaluated. The author suggests a framework of seven categories of evaluation questions which are (p. 134): program design and implementation; learner participation; learner satisfaction; learner knowledge, skills, and attitudes; application of learning after the program; the impact of application of learning; and program characteristics associated with outcomes.

In an earlier study, Cervero and Rottet (1984) found that there are other important variables that determine the effectiveness of CPE programs. The evaluation model using the one-variable approach is not useful to adult educators for improving future CPE programs. As a result, Cervero developed a conceptual model using four sets of independent variables that can be used for program evaluation. The components include the characteristics of (1) individual professional; (2) the CPE program; (3) the nature of proposed changes; and (4) the organizational social system (Cervero, 1985). To be more useful, evaluation should be more comprehensive, analyzing the relevant variables influencing the effectiveness of educational programs.

Other examples of adult education program evaluation approaches found in the literature are suggested by Knowles (1970), Houle (1980), and Knox (1986). According to Knowles (1970), evaluation in adult education has been overemphasized to the extent of underproduction
of "practical, feasible, and artistic evaluations in terms of program review and improvement" (p. 219). The author recommends that adult educators should practice evaluation congruent with the philosophy of adult education field. Knowles (1970) suggests four simple but comprehensive steps of program evaluation (p. 223). They are (1) formulating questions to be answered (or establishing the criteria, yardsticks, or benchmarks), (2) collecting appropriate data to answer above questions, (3) analyzing and interpreting data as answers to evaluation questions, and (4) modifying and improving programs based on the findings.

Similarly, Houle (1980) also suggests a comprehensive framework of evaluating CPE programs. The author suggests the measurement of program results, the assessment of the quality of practices as a result of educational activities, and a general measurement of impact on the entire profession. On the other hand, Knox (1986) suggests that effective evaluation should be designed in the context of program development. The author suggest the use of a formal evaluation to accurately describe influences, performance, and expectation; to make sound judgements based on empirical evidence; and to communicate findings which will encourage stakeholders to improve future program. In general, program evaluation suggested by these authors seems to advocate the use of a formal, systematic, and comprehensive approach.

In summary, there are various types of theories and models in program evaluation which have been constructed and reconstructed over time. Some of them are quite different from one another but some overlap. However, many of the models advocate the use of evaluation methodology which resemble a research endeavor. Since evaluation is an applied research, the models advocate the use of a formal, systematic, and comprehensive approach to data gathering and analysis regardless of which paradigm is being used. This is to ensure pertinent and defensible evidence of evaluation findings.
The Practice of Program Evaluation

One of the major dilemmas facing the field of evaluation is in the arena of evaluation practices. In general, the everyday practice of program evaluation is much simpler than the prescription given in the literature. Evaluation practices are criticized for not doing what is prescribed in evaluation theories and models. There are ample amounts of research in the practice of program evaluation. However, the scope of most existing research focuses only on the methodology of evaluation practices (Shadish and Epstein, 1987).

According to Stufflebeam and Shinkfield (1985), the practice of evaluation "traditionally has been at best uneven and, in general, poor" (p. 2). The authors further stress that professionals must continuously obtain pertinent evaluative feedback to satisfy the needs of the stakeholder. The evaluation process must include the need assessment, assessing the plan of intervention, closely monitoring the program implementation, assessing short- and long-term outcomes, and searching for more effective and efficient ways of providing the services. Rossi and Freeman (1993) also criticize the practice of impact evaluations. According to the authors, normally the practices are not rigorous and, as a result, the findings are not complete and are limited in use. The authors suggest that in order to achieve a higher degree of plausibility, impact assessment has to adhere to rigorous research procedures.

Shadish and Epstein (1987) conducted a study to look for patterns of program evaluation practices among members of the Evaluation Research Society and Evaluation Network. Among other findings, the study shows that only eight percent of the samples were trained specifically as evaluators. The majority of the samples learned evaluation directly from their own working experience, from reading books and articles, and from discussions with other evaluators. Another interesting finding is that the sample has a low level of familiarity with evaluation theories and
concepts described in the literature. Of 21 influential evaluation works in the literature, 71 percent were not fully recognized by the majority of the sample. These include all the work of Stake, Scriven, and Wholey. Whereas the nonrecognition rate for the Cronbach's classic work on evaluation for course improvement is 72 percent, the rate for Campbell's comments on case studies is 63 percent, and for Chen and Rossi's theory-driven evaluation is 83 percent. These findings lead to the authors' concern about the danger of scholarly illiteracy in evaluation. The authors suggest that the illiteracy problem is probably due to the failure of evaluators to read the literature sufficiently or to the incompatibility of theory to practice.

In a study to examine evaluation methods in business-education partnerships, Erickson (1991) finds that there is little standardized evaluation methodology being practiced. The evaluation activities are inconsistent, loosely structured, rendering superficial, impressionistic judgments of program effectiveness. In another study, Shireman (1991) analyzed how evaluations are utilized by hospital-based health education programs that are patient/client centered. Among other things, the findings suggest that program evaluation is not widely used and it is limited in scope. The primary source of evaluation data comes from the providers rather than the program recipients. Patients or clients are not being involved in the planning and implementing of program evaluation.

In yet another study, Lynch (1988) analyzes evaluation practices of educational programs reviewed by the Joint Dissemination Review Panel from 1980 to 1983. The study finds that "fewer than half of the evaluations were high enough quality to produce reasonably believable evidence of program effectiveness" (p.273).
The Practice of Evaluation in Adult Education

Similarly, the practice of evaluation in adult and continuing education receives much criticism in the field's literature. As stated by Sork and Caffarella (1989), many authors have argued the value of program evaluation in adult and continuing education field. Adult educators do not seem serious in evaluating their programs. "The action of practitioners appears to be in accord with the sentiment that evaluation of adult education programs is either not important or not feasible" (p.242). As noted by Knox (1986), the majority of problems in program evaluation are due to the inadequate planning of data analysis procedures. As a result, little relevant data are collected.

Since evaluation is an important part of the program development process, any program development research is bound to have some element of program evaluation. Pennington and Green (1976) conducted a study to analyze the program development process in six different professions. The authors find that the term comprehensive evaluation "does not describe evaluation practices in continuing professional education program as observed in this study" (p. 21). The practice of impact assessment is superficial and only utilizes one or a few out of many methods suggested by the literature. The authors suggest that evaluation should be viewed as a long-term investment, and adult educators should be aware of the existence of various cost-effective evaluation methods.

In another study of agricultural extension agents, Casey (1989) found that "many agents said they didn't do formal evaluation but relied on informal evaluation" (p.101). The respondents were not very positive about evaluation, even though they knew that they had to do it. The findings of the study show that the agents have "lukewarm" attitudes toward evaluations. In addition, the respondents view evaluations as a burden put on them by supervisors and the
organization. Some respondents see no benefit of conducting formal evaluation. They see evaluations as things that are required only by their supervisors and the organization and not by them.

There are only a few studies specifically examining the practice of program evaluation in adult and continuing education field. The first study was conducted by Imel (1976) to explore information about the practice of program evaluation in postsecondary adult and continuing education organizations. The respondents were public community and four-year colleges and universities in five states. According to Imel, in 1976, the field of adult and continuing education had not yet developed any systematic evaluation. No indication of systematic evaluation was found in the literature. As a result, the author claims that adult educators do not practice systematic evaluation. Nevertheless, one of the author's findings shows that adult educators do practice systematic evaluation. However, Imel only defines systematic evaluation as having administrative provision for evaluation and having an evaluation plan of some kind.

On the other hand, Frazer-Koontz (1991) conducted a study to examine the attitudes, competence, and practices of human resource directors toward their training-program evaluations. The study's sample was selected from 150 of the largest banks in the United States. Although the questionnaire's return rate is low (38 percent), the findings are worth looking at in comparison with other studies. The results show the respondents have positive attitudes about training-program evaluations. They always support the evaluation program. However, a majority of them only used reaction evaluations to judge program effectiveness, even though they knew it was not really useful. The directors also rate themselves as less competent in planning and implementing impact assessment.

Two other studies of evaluation practices in continuing education were carried out by
Brady (1986) and Gutek (1988). Brady's study examines whether any program evaluation model is being used in the evaluation practice of baccalaureate nursing schools throughout the country. She concludes that the practice does not take advantage of using any evaluation model. Therefore, the practice is significantly different from what is prescribed in the models. Baccalaureate schools of nursing tend to combine the program evaluation process with criteria for accreditation. On the other hand, Gutek's (1988) study was to investigate the perception and discrepancy between program evaluations described in the literature with training-program evaluation practices in non-manufacturing business organizations. The findings show that there is a positive perception of the value, needs, and feasibility of training-program evaluation by training directors with the frequency of conducting evaluation activities. However, the need of conducting evaluation is very low, and there is no demand on the part of the organization to evaluate training programs.

In summary, the practice of program evaluation is considered to be different compared to the theory and models prescribed in the literature. Most of the time, the practices are informal, unsystematic, and not comprehensive. Many of the empirical studies supporting this statement focus on the methodology of evaluation. The studies examined how practitioners conducted the process of evaluation in the field. For example, the studies asked what types of evaluation design were being used, at what point in the program planning process the evaluation was conducted, what was the source of data, how were evaluation findings utilized, and the like. This is not surprising since evaluation design and methodology are very important aspects of conducting effective evaluation.
**Suggested Factors Influencing Evaluation Practices**

The complexity and the contextual nature of the evaluation process lends itself to be susceptible to many influencing factors. The quality of evaluation depends on the suitability of the approach used, the methodology, the implementation, and the utilization of findings. This section will try to gather any suggested factors which influence evaluation practices. Most of the factors are suggested based on opinion and experiences, and some are suggested indirectly through research which was not designed to investigate the factors.

Rossi and Freeman (1993) suggested a set of common constraints and limitations of program evaluation which include politics, ethics, time, money, and human resources. Among these factors, politics is considered to be the most influential factor of program evaluation. Other authors who mentioned politics are Cronbach et al. (1980), Smith (1981), Palumbo (1987), Mohr (1988), Hendrick (1988), and Alkin (1990). The ethical factor is also mentioned by Hendrick (1988) and Newman and Brown (1992). Lack of ethics on the part of evaluators or organizations could deviate evaluation from serving the stakeholder into a selfserving mode. Doing pseudoevaluation such as covert study is one example of an unethical evaluation practice. Whereas money is mentioned by Larson (1985), money and time are also pointed by Reinecke and Welch (1986). A formal, systematic, and comprehensive approach of program evaluation requires a considerably large investment of money, time, and other resources. Limited money and time might shortcut evaluation, and as a result the finding would be less useful.

In a study to describe evaluation practices, Shadish and Epstein (1987) suggest some human factors involved in determining the differences between theory and practice. The authors suggest that evaluators' knowledge, skills, experiences, and preferred theory have a major impact on evaluation practices. Evaluators' and administrators' perceptions and attitude are also appeared
to influence practice (Gutek, 1988; Casey, 1989; Frazer-Kontz, 1991) as does evaluators' supervision (Palmisano, 1981), conflict between evaluators and organization's priorities, and human resistance to evaluation (Posavac and Carey, 1985).

Other important suggested factors affecting evaluation practices are school's policy (Sussman, 1991) and work setting demand (Shadish and Epstein 1987). On the other hand, Wilcox (1989) finds four salient factors determine how evaluation is practiced in schools setting. They are participation of stakeholder, the role of evaluators, the purpose of evaluation, and the identification of evaluation criteria.

According to Stake (1981), the peculiarity of the adult and continuing education field also contributes to the gap between theory and practice of program evaluation. The "nature of clientele and (the) nature of programs, therefore, can contribute to the 'evaluation attitudes' of adult educators" (p. 4). The practice is the reflection of the field's philosophies, clientele, and programming thrusts. In addition, the humanistic orientation and the voluntary nature of the clientele "have contributed to a nonscientific attitude about evaluation processes" (p.23).

In sum, the factors suggested to influence evaluation practices make the practices dynamic and idiosyncratic. There is a possibility that those factors are not always present in every practice. Furthermore most of the factors are just suggested or perceived. It is not being investigated in detail in an empirical manner. The complexity of the evaluation process and the need of evaluators to be on top, make the examination of the influencing factors even more complicated. Nevertheless, there is an emergence need to develop theory of practice which takes these factors into consideration especially in the field of adult and continuing education.
Evaluation and Adult Education in Malaysia

Program evaluation has become a major concern in planning and implementing development programs in Malaysia. Since a vast amount of resources has been committed to the programs, accountability becomes the impetus for the rise of concern regarding program evaluation. The growth and development of the evaluation field in the West, especially in the United States, has been a major influence in Malaysia.

The importance of evaluation in Malaysia can be seen in the establishment of evaluation agencies and units in public and private organizations. The Malaysian government has mandated that every public funding program be evaluated (Maimunah, 1990). There are two federal agencies responsible for evaluation work. One of them is INSPEK (Institut Penilaian Negara) or National Institute of Evaluation. This agency is responsible for evaluating nearly all national programs including programs in the adult and continuing education field. Another federal agency dealing with evaluation is the Evaluation Unit at the Prime Minister Department. This unit is responsible for evaluating governmental special projects and programs. The Social Economic Research Unit (SERU) is another government agency having an interest in program evaluation. As the name implies, SERU conducts evaluation research on socioeconomic programs when the need arises. In addition, every government ministry and department has an evaluation unit of some sort (Maimunah, 1990).

The importance of evaluation in Malaysia is also shown in several studies of extension workers' training needs. In one such study of professional workers in the rubber industry, Shamsudin (1977) found that program evaluation is one of the important areas where training is needed. Saidin (1981) found that program evaluation is rated fourth in terms of the importance of extension supervisors' training needs. The author also found that the importance of training needs
is significantly related to the required professional expertise.

Maimunah (1990) advocates the use of a formal, systematic, and comprehensive approach of program evaluation in Malaysia. According to the author, formal evaluation uses the scientific method which relies on systematic data collection and data analysis. Formal evaluation also should adhere to scientific research procedures. Maimunah (1990) recommends the use of the evaluation model suggested by Stufflebeam, Frutchey, Bennet, and Steele. According to Maimunah (1990), Frutchey's model asserts the use of formal evaluation through the three decision making processes of developing evaluation criteria, collecting evidence, and making judgments. Bennett (1976), as described by Maimunah (1990), suggests the use of several criteria. The stepwise criteria are impact, activity, involvement, reaction, change in PSKA (practice, skill, knowledge, and aspiration), and change in impact. Maimunah (1990) also suggests Steele's concept of criteria in explaining the discrepancy evaluation model. According to Maimunah (1990), systematic evaluation identifies criteria in advance. These criteria will indicate what information and evidence to collect and from where. Then what is actually observed to be the result of educational programs will be compared to the criteria for making evaluation judgements. Maimunah (1990) also suggests the use of Stufflebeam's CIPP model.

The recommendation to use a systematic and comprehensive evaluation approach is also found in a study by Sulaiman, Rahim, Azimi, and Musa (1985). The authors suggest some areas where quantitative indicators of success should be the basis for evaluating community development programs for Malaysian fishermen. Some of the suggested areas are health education, group actions, respondents' inspirations and perceptions, supplemental income activities, etc. In addition, the Center of Extension and Continuing Education at the Universiti Pertanian Malaysia also advocates the use of systematic and comprehensive evaluation through
its academic and inservice training courses. The center also produces some training modules to be used in the training of development workers (Sulaiman, Saidin, and Asmah, 1986).

In practice, program evaluation is not systematically conducted in Malaysia. According to Maimunah (1990) most of the evaluation used by extension agencies are ad hoc and informal. As a result the validity of the findings are questionable and not comprehensive.

In sum, the preferred evaluation approach in Malaysia is comparable with the preferred paradigm in the United States. The ideal evaluation approach is the use of a formal, systematic, and comprehensive evaluation. However, there is an indication that evaluation practices in Malaysia do not follow the theory and model prescribed in the literature. This phenomena is comparable to what is happening in the United States. There is a disjuncture between theory and practice of program evaluation.
BIBLIOGRAPHY


Training evaluation: clients' roles

Junaidah Hashim

The Authors

Junaidah Hashim, Department of Business Administration, Kulliyah of Economics and Management Sciences, International Islamic University Malaysia, Kuala Lumpur, Malaysia

Abstract

Training evaluation is an elusive concept, especially when it comes to practice. The practice of evaluation in training has received a lot of criticism. This criticism is largely explained by the unsystematic, informal, and ad hoc evaluation that has been conducted by training institutions. In Malaysia, training activities are monitored by the government. Organisations are required to obtain training services from approved training providers registered with the government. Examines the clients' demand toward evaluation, the commitment given by training providers, and the overall practice of evaluation by the training providers in Malaysia. Finds that the government, client, and economic situations have influenced the evaluation practice in a positive direction.

Article type: Survey.

Keywords: Training, Evaluation, Malaysia, Commitment, Clients.

Content Indicators: Research Implications** Practice Implications** Originality** Readability**

Introduction

This paper addresses the issues of training evaluation practices in general, and examines the training evaluation in Malaysia through a case study.

The Malaysian government since her independence in 1957 has manifested her commitment toward education and human resource development. The emphasis was on education, because the government believed that it was the key input to national development. The government has recognised the importance of human resource development in its quest for achieving developed nation status. This commitment was translated into the establishment and growth of training agencies in the country.

In Malaysia, as of 1999, records show that it has about 300 training institutions registered as training providers to various companies in the country. For a country like Malaysia, this number is considered large and it has made the training industry very competitive. The establishment of these institutions may have resulted from the government's new legislation, which requires every company to promote training for its workforce to ensure that it is competent and can further contribute to the country in realisation of Vision 2020. For that, the Malaysian government has passed an Act of Parliament entitled the Human Resource Development Act 1992, which requires companies to contribute a 1 per cent equivalent of its monthly payroll to a fund which would then be used to promote training. A special council was set up to manage this fund and to monitor the training activities for private companies in Malaysia. To facilitate the monitoring activities, companies are advised to obtain training services required from the council registered training providers only. This has caused the number of training providers in Malaysia to increase rapidly to cater for the demand.
Case study

Malaysia has a vision to be a fully developed nation by the year 2020. To achieve this vision, Malaysia needs a highly educated and trained workforce. This effort is shared by both the public and the private sectors by not only allocating greater budget for training but also by making a contribution to the Human Resource Development Fund. Since a vast amount of resources has been committed to training programmes, and with the prevailing situation of economic constraint, the demand for justifying training expenses is gaining impetus. Being an agency responsible for human resource development and training for the private sector in Malaysia, the Human Resources Development Council is aware about the training programme needs. For this purpose, it has designed a standard evaluation form and it requires all training providers to fill in this form and return it to the Council for further analysis.

The author conducted a research project to find out what the practices of training providers in Malaysia are. The study covered all the training providers, who were registered with the Human Resource Development Council as approved training providers. The actual sample for this study was 262 institutions. The response rate was 49 per cent.

This study used a specially constructed questionnaire with a five-point Likert-scale (Cronbach’s coefficient alpha = 0.7310 during pre-testing). In the first part of the questionnaire, respondents were asked to indicate how frequently each of the evaluation methods listed were used in the evaluation they conducted.

The results show that the respondents used all the evaluation methods commonly found in the literature. Trainee feedback was the most frequently used evaluation method (mean 4.19) by the respondents, as shown in Table I. Besides trainee feedback, other frequently used methods were observation (mean 3.84), interview (mean 3.65), performance analysis (mean 3.56) and reaction form (mean 3.49).

The second part includes questions to find out how the evaluation was planned, and how frequently the evaluation was carried out. Stufflebeam (1985) suggested that evaluation should consist of context, input, processes, and product evaluation. These evaluations are done at different stages during the programme development.

Table II shows the percentage of responses to each statement asked. Respondents agreed that they evaluate their training right after the training is completed (88.3 per cent and mean 4.49). Evaluation during the implementation was the second agreed by the respondents (mean 3.79). Some respondents did evaluate their training during the planning stage (mean 3.57). Each statement relevant to formal, comprehensive and systematic evaluation was then computed as a total score and the results are depicted in Table III. It indicates that respondents agreed to some extent conduct formal, comprehensive and systematic evaluation.

Much of the literature has highlighted that clients seemed not to demand that the training providers conduct evaluation for the training they provided (Smith and Piper, 1990). This study attempted to find out some information about clients’ demand towards evaluation from the training providers’ perspective. Eight statements initially contained in the questionnaire, but three of them were deleted due to their reliability. The five statements asked and the responses received from respondents were depicted in Table IV.

Referring to Table IV, the item that has the highest mean was "clients require reaction evaluation" (mean 3.71). The responses for other items were relatively high, all the means being above 3.00. The five statements were then computed as a total score, for the purpose of examining the overall demand from clients. On the overall score, the clients’ demand was moderate (3.38). In the respondents’ opinion, clients did to some extent demand that they conduct evaluation.

This study involved corporate clients who hired training providers from the Human Resource Development Council, and it is found that clients definitely preferred a training package that includes evaluation, because they pay for the training. Training providers have to oblige the clients’ requirement, if they want to provide the training services to these clients and plan to continue doing so.

The other part of the questionnaire attempted to find out training institutions’ commitment towards evaluation. Eight items were asked for this purpose (Cronbach coefficient alpha = 0.8089). It was found that training providers’ commitment was moderate (mean 3.05). More than half of the respondents strongly agreed (66.0 per cent) that effective training ensures improved performance (mean 4.44). Most of the respondents (80.9 per cent) strongly disagreed (mean 1.30) that conducting evaluation is a waste of time. The eight statements asked were computed as a total score. Overall commitment as shown in Table V.

http://alidoro.emerald-library.com/vl=7474185/cl=30/nw=1/rpsv/cw/mcb/030905.../p374.htm 4/5/02
### Table I
Distribution of respondents by evaluation method

<table>
<thead>
<tr>
<th>Methods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey</td>
<td>6.4</td>
<td>10.6</td>
<td>34.6</td>
<td>26.6</td>
<td>22.3</td>
<td>3.48</td>
</tr>
<tr>
<td>Interview</td>
<td>3.2</td>
<td>8.5</td>
<td>26.6</td>
<td>43.6</td>
<td>18.1</td>
<td>3.65</td>
</tr>
<tr>
<td>Observation</td>
<td>1.1</td>
<td>7.4</td>
<td>21.3</td>
<td>46.8</td>
<td>18.4</td>
<td>3.84</td>
</tr>
<tr>
<td>Document review</td>
<td>4.3</td>
<td>18.1</td>
<td>31.9</td>
<td>30.9</td>
<td>14.9</td>
<td>3.24</td>
</tr>
<tr>
<td>Organizational audit</td>
<td>22.3</td>
<td>20.2</td>
<td>31.9</td>
<td>19.1</td>
<td>6.4</td>
<td>2.67</td>
</tr>
<tr>
<td>Performance analysis</td>
<td>6.4</td>
<td>10.6</td>
<td>29.8</td>
<td>26.6</td>
<td>26.6</td>
<td>3.56</td>
</tr>
<tr>
<td>Export review</td>
<td>19.1</td>
<td>17.0</td>
<td>37.2</td>
<td>19.1</td>
<td>7.4</td>
<td>2.79</td>
</tr>
<tr>
<td>Panel checklist</td>
<td>28.7</td>
<td>14.9</td>
<td>24.5</td>
<td>22.4</td>
<td>8.5</td>
<td>2.08</td>
</tr>
<tr>
<td>Site visit</td>
<td>11.7</td>
<td>21.3</td>
<td>26.6</td>
<td>29.8</td>
<td>10.6</td>
<td>3.06</td>
</tr>
<tr>
<td>Pilot tests</td>
<td>28.7</td>
<td>11.7</td>
<td>25.5</td>
<td>26.6</td>
<td>7.4</td>
<td>2.72</td>
</tr>
<tr>
<td>Training feedback</td>
<td>5.3</td>
<td>11.1</td>
<td>12.8</td>
<td>30.9</td>
<td>50.0</td>
<td>4.19</td>
</tr>
<tr>
<td>Simulation</td>
<td>23.4</td>
<td>22.3</td>
<td>26.6</td>
<td>19.1</td>
<td>8.5</td>
<td>3.67</td>
</tr>
<tr>
<td>Self-report</td>
<td>10.6</td>
<td>22.3</td>
<td>35.1</td>
<td>21.3</td>
<td>10.6</td>
<td>2.99</td>
</tr>
<tr>
<td>Work samples</td>
<td>18.1</td>
<td>19.1</td>
<td>25.5</td>
<td>27.7</td>
<td>9.6</td>
<td>2.91</td>
</tr>
<tr>
<td>Peer report</td>
<td>14.9</td>
<td>26.6</td>
<td>35.1</td>
<td>18.4</td>
<td>2.3</td>
<td>2.72</td>
</tr>
<tr>
<td>Supervisor</td>
<td>12.8</td>
<td>14.9</td>
<td>34.0</td>
<td>27.7</td>
<td>10.6</td>
<td>3.08</td>
</tr>
<tr>
<td>Competency test</td>
<td>18.1</td>
<td>12.8</td>
<td>28.7</td>
<td>23.4</td>
<td>17.0</td>
<td>3.08</td>
</tr>
<tr>
<td>Reaction form</td>
<td>11.7</td>
<td>8.5</td>
<td>28.7</td>
<td>21.3</td>
<td>29.8</td>
<td>3.49</td>
</tr>
</tbody>
</table>

Notes: 1 = not at all, 2 = seldom, 3 = sometimes, 4 = frequently, 5 = very frequently; N = 94

### Table II
Distribution of respondents by evaluation schedule

<table>
<thead>
<tr>
<th>Formal evaluation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We evaluate our training programme before we plan for the training</td>
<td>17.0</td>
<td>7.4</td>
<td>16.0</td>
<td>33.0</td>
<td>26.6</td>
<td>3.45</td>
</tr>
<tr>
<td>We evaluate our training programme during the planning stage</td>
<td>13.8</td>
<td>3.2</td>
<td>12.8</td>
<td>52.1</td>
<td>18.1</td>
<td>3.57</td>
</tr>
<tr>
<td>We evaluate our training programme during the implementation stage</td>
<td>13.8</td>
<td>5.3</td>
<td>3.5</td>
<td>41.5</td>
<td>30.9</td>
<td>3.70</td>
</tr>
<tr>
<td>We evaluate our training programme right after the training is completed</td>
<td>3.2</td>
<td>3.2</td>
<td>3.2</td>
<td>18.1</td>
<td>70.2</td>
<td>4.49</td>
</tr>
<tr>
<td>We evaluate our training programme sometimes after the training</td>
<td>14.9</td>
<td>17.0</td>
<td>12.8</td>
<td>39.2</td>
<td>19.1</td>
<td>3.28</td>
</tr>
<tr>
<td>Systematic:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>We evaluate all of our training programmes</td>
<td>2.1</td>
<td>10.6</td>
<td>8.3</td>
<td>25.5</td>
<td>53.2</td>
<td>4.17</td>
</tr>
<tr>
<td>Not all of the training programmes need evaluation</td>
<td>10.6</td>
<td>30.9</td>
<td>13.8</td>
<td>13.8</td>
<td>30.9</td>
<td>3.23</td>
</tr>
<tr>
<td>We have a regular evaluation schedule for certain programmes</td>
<td>6.4</td>
<td>10.6</td>
<td>12.8</td>
<td>40.4</td>
<td>29.8</td>
<td>3.77</td>
</tr>
</tbody>
</table>

Notes: 1 = strongly disagree, 2 = mildly disagree, 3 = neither disagree nor agree, 4 = mildly agree 5 = strongly agree; N = 94

### Table III
Summary of statistics of respondents by comprehensive, formal and systematic evaluation

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal</td>
<td>3.71</td>
<td>0.6970</td>
</tr>
</tbody>
</table>


Shamsuddin A., 1995, "Contextual factors associated with evaluation practices of selected adult and continuing education providers in Malaysia", PhD dissertation, University of Georgia, Athens, GA.


CONTEXTUAL FACTORS ASSOCIATED WITH EVALUATION PRACTICES
OF SELECTED ADULT AND CONTINUING EDUCATION
PROVIDERS IN MALAYSIA

Shamsuddin Ahmad, Ed. D

The author is a Lecturer at The Center for Extension and Continuing Education, Universiti Pertanian Malaysia. He graduated with a doctoral degree from The University of Georgia, Athens in Spring 1995. Specializations: program planning, evaluation, and youth development.

Abstract: Provided in this qualitative multiple case study are descriptions of how selected developmental agencies in Malaysia carried out program evaluation and what factors influenced the practice. Data were gathered from interviews with evaluation practitioners and from official documents. Evaluation practice of the agencies was not systematic nor comprehensive. Four contextual factors shaping the practice that emerged from the data are discussed.

There are more than fifty different theories and models of program evaluation found in the literature (Steele, 1989). These theories and models are general in nature and are prescribed to be applied in any situation. They are generally designed to be context-free and do not include suggestion for situational adaptions. On the other hand, evaluation practices are context specific. The practices are carried out in many different contextual environments. In deciding which evaluation approach to use, evaluation practitioners must consider many contextual factors (Shadish & Epstein, 1987) ranging from the political and organizational to personal milieu. This idiosyncratic nature may explain why evaluation practices are different from approaches prescribed by theory and models in the literature.

The reason of the discrepancy between program evaluation theory and practice is not well-researched, especially from the perspective of the practitioners of evaluation. Previous studies on the practice of program evaluation have not investigated in detail the contextual factors that influence how an evaluation is implemented. Most of the studies focus on the comparison of actual evaluation practices with that of the method suggested by evaluation models (Shadish & Epstein, 1987). Some authors suggest the lack of resources as the main obstacle (Sork & Caffarella, 1989), while others suggest inadequate planning for conducting the data analysis (Knox, 1986). In addition, Rossi and Freeman (1993) suggest that politics, ethics, time, money, and availability of human resources are the constraints and limitations that prevent the use of a more formal, systematic, and comprehensive approach to program evaluation. On the other hand, Forester (1976), Knowles (1980), Brookfield (1986), and Steele (1989) question the existing evaluation theories and models prescribed in the literature. According to them, many models are not suitable for the adult and continuing education field. The models do not take into consideration the nature and philosophy of adult and continuing education.

The preferred evaluation approach in Malaysia is comparable to the preferred paradigm in the United States. The ideal evaluation approach suggested in Malaysian literature is the use
of a formal, systematic, and comprehensive evaluation (Maimunah, 1990). However, it is suggested that the practice of program evaluation in Malaysia also do not follow the theory and models prescribed in the literature. This phenomena is comparable to what is happening in the United States. There is a disjunction between theory and practice of program evaluation and there is no specific study focusing on what and how contextual factors influence the practice of program evaluation. Therefore, the purpose of the study was to identify how contextual factors shape evaluation practices. Specifically, the study addressed the following questions: (1) What are the evaluation practices of different adult and continuing education agencies in Malaysia? and (2) What are the contextual factors involved in the evaluation practices of different adult and continuing education agencies in Malaysia?

METHODOLOGY

This study used the multiple qualitative case study approach because it provided more compelling and robust findings (Yin, 1989); it allows comparison of evaluation practices from different kinds of agencies, and it gives some variation of cases being studied. In addition, it allows an in-depth, holistic description and interpretation of complex real-life phenomena (Merriam & Simpson, 1989). In this manner, the research questions for this study were answered with more depth, concreteness, and more contextuality.

Criteria for Sample Selection

This study used purposive sampling to select agencies which provided a variation that would enrich the findings of the study. Three agencies were selected to represent the major program areas: agricultural development, entrepreneurial development, and continuing education. These three program areas were chosen because they were important both to the development of the country and to the adult and continuing education field in Malaysia (Malaysia, 1986; Universiti Pertanian Malaysia, 1983; Bahman, 1992). These samples provided a maximum variation of sample suitable for the multiple case approach, which represented a range of a particular dimension (Patton, 1990). The sample selection also would give a variation for cross-case analysis.

The selection of the three agencies was determined by recommendation from one or four local adult education professors. A set of criteria was given to the panel as a guideline for selecting the respective agencies. The criteria included agencies which: (1) Conduct program evaluation on a regular basis (2) Have written policies or procedures on evaluation (3) Have assigned personnel responsible for evaluation and (4) Are perceived by the panel members to have successful programs.

Data Collection

This study triangulated data from interviews with evaluation practitioners and official evaluation documents. A semi-structured interview guide was used as a checklist to make sure the issues and questions were explored and asked during the interview. Since the interview was conducted in the Malaysian Language, the interview guide was translated before any actual interview was conducted. The translation was checked and verified by a certified translator in Malaysia. All interview were audio recorded with the respondents' permissions.

Evaluation documents collected in the study provided a method of verifying and
augmenting the information provided through the interviews. Such documents, which are products of their context, provide real evidence grounded in real practice (Merriam, 1988). There was a variety of documents related to the purpose of this study provided by the evaluators. For example annual agency report, booklets and brochures about the agency programs, evaluation questionnaires, evaluation reports, and evaluation guide.

Data Analysis

Two types of data analysis were carried out in this study. First, data was analyzed within each case followed by cross-case analysis. Cross-case analysis was conducted to build “categories, themes, or typologies that conceptualize the data from all the cases” (Merriam, 1988, p. 155). Basically both types of analysis were the same except that cross-case analysis was carried out to compare the initial findings from each of three cases. Cross-case analysis is useful to build a general explanation that is common for all the three cases (Yin, 1989).

The data were analyzed simultaneously during the data collection phases. Pre- or initial analysis was started after one interview had been completed and transcribed in order to discover any relationships, patterns, constructs, or themes within the information. This initial analysis was used to refine future interview and document analysis. In the initial analysis, data were categorized, refined, fleshed out, changed or challenged. This process was continued until sources were exhausted and the category was saturated.

FINDINGS

The findings are divided into two sections: First, the case description of evaluation practices of each agency, and second, a cross-case analysis of contextual factors which shape evaluation practices in all three agencies.

Case Descriptions

Agency A was selected to represent continuing education programs. With a mission of “developing the nation through excellence in training,” this agency is responsible for training public employees in the broad areas of administration and management. There are three types of evaluations conducted in this agency which include formal and informal approaches. However, the only formal evaluation required by the management is the end-of-training evaluation using the “pink form.” The mid-evaluation and the oral evaluation which are initiated by the facilitator and the coordinators are informal. These two informal evaluation findings are not reported in the final quarterly and yearly evaluation reports which are sent to the agency’s management.

Agency B, a statutory body, has offices in every state and district in the country, and is one of the main agricultural agencies in Malaysia. It is mandated to develop the smallholders’ rubber plantation with less than 100 acres of rubber trees. The agency’s mission is “to establish a prosperous smallholders’ community through the development of every aspect of social-economic endeavor.” Program evaluation in Agency B is conducted through regular meetings, periodic evaluation, and “on request” activity. The periodic evaluation activities which are required by the agency’s administration provide information mainly on the amount of program resources that have been used, the number of clientele, and the size of rubber plantation covered by the programs. Data for this type of evaluation are gathered continuously throughout the programs. However, the latter evaluation which originated from demand outside of the agency
examines the effectiveness as well as the impact of the programs while the programs are running. By doing “on request” evaluation projects with outside help, program evaluators in this agency gain experience in planning and implementing systematic program evaluations.

Agency C was selected to represent agencies which offer educational entrepreneurship programs to their clientele. According to the agency’s yearly report, the mission of this agency is to “upgrade the economic and social status of ‘Bumiputera’ (indigenous people) through activities in entrepreneurship, corporations, equity ownership, and human resources development.” Except for the entrepreneurship training program, which also measured program effectiveness by using end-evaluation questionnaires, program evaluation in Agency C was focused more on reporting program implementation as a measurement of success. Most of these evaluations were made for the purpose of budgeting and reporting the amount of money spent and how many projects or clientele had been served. There were concerns by the respondents wanting to do more systematic and comprehensive evaluations. However, without the directive from the management, the practice would stay as it was.

**Contextual Factors Involved in Evaluation Practices**

There were four categories of contextual factors found in the data collected. All four factors were involved in the evaluation practice of all three selected agencies. All the factors were closely related and intertwined with at least one other factor. Together as a group, the factors shaped the evaluation practice of the selected agencies.

Under the first factor, which was the importance of evaluation to management there were three elements that shaped evaluation practice: (1) Mission of public organizations (2) Organizational structure of the organization and (3) Management’s attitude towards evaluation. The agency mission and how it was conducted through the nature of the programs made systematic and comprehensive evaluations inapplicable or unnecessary. The highest priority of the mission was to “help” and “give” the programs to the people. Evaluators were also directed by the management to do some form of evaluations. These evaluations were to determine the “success” of a program by showing how much input had been spent and how many people had received the program. The management only perceived evaluation from a narrow perspective and did not think in-depths evaluations were important.

Abilities of evaluators also influenced how evaluation was practiced. Their perceptions of evaluation were limited to superficial levels of objective-oriented and “strength-and-weaknesses” approaches to evaluation. Many evaluators were not fully equipped with proper program evaluation knowledge and skills because they were not perceived to be needed by the management. Evaluators were not receiving any specific evaluation training except for a few who attended short evaluation courses or had evaluation courses while studying at a university.

Culture of the clientele also influenced how evaluation was practiced. By nature, the agencies’ clientele were not critical when evaluating programs. As a result, evaluation was usually high and thus gave the impression that the program and the evaluation were appropriate. Naturally the evaluation practice was accepted and continually used. The clientele’s behavior also was a form of paying back the favor for programs and services they received from the agencies.

The last contextual factor was the tradition-bound evaluation practice. Evaluators were most likely to do evaluation according to what had been done before by them or by previous
evaluators. This practice had become a ritual tradition that was difficult to change. Evaluation practice was taken for granted and was never formally discussed in the agencies.

SUMMARY AND DISCUSSIONS

Even though program evaluation was officially conducted in each of the agencies, the approach was minimal. The results were mainly used to “show” program success and do fiscal budgeting. Data collection consisted of a basic level of information such as the number of program inputs that had been used and the reactions of the clientele toward the programs. The practice was limited to the use of simple standardized forms and questionnaires to be completed by the evaluators and the clientele. Data were analyzed using only simple statistics such as mean, percentage, ranking, and rating. There was no evidence of any known theory or model being used. As one evaluator phrased it, “I’m not familiar with the theory and I don’t know how to answer that; we just follow what our seniors did.”

The findings also revealed that the four contextual factors involved were interconnected in shaping the evaluation practice. They were very specific in nature and specific in their influence on the evaluation practice. Collectively, these factors shape evaluation practice by encouraging the minimalist approach to program evaluation.

The conclusion that the evaluation approach practiced by the agencies was not systematic and comprehensive supports pervasive research on evaluation practiced conducted in the United States. Even though Tyler’s objective-oriented evaluation approach is the most prominent concept used by many practitioners, the practices were much simpler. Tyler’s model is more systematic and comprehensive which requires a set of specific steps to be taken in order to develop the learner objectives and the appropriate evaluation instrument. Then the objectives are tested using a pre- and post-research design to find out the congruence between the objectives and the learner outcomes. Evaluation was also perceived and used to improve the program by determining the program’s “strengths and weaknesses.” “We do this evaluation to look at what is good and bad about the programs” and “We ask what their [the clientele] reactions to the programs are, both oral and written.” This perception of evaluation was only targeted at the lowest level of an evaluation hierarchy.

While the previous research literature suggests a more general group of factors, this study found four specific factors in a group that co-exists with one another in shaping the evaluation practice. The four contextual factors were more explicit, complex, and broader in the sense that several factors were influencing evaluation practice. These factors which were identified from the study corroborate and expand the list of contextual factors which are mainly suspected or suggested in previous literature.

Evaluation practice is a complex activity, and it is not that easy to follow existing theory or models because the factors influencing the practices are many and connected. Indeed, evaluators are faced with a situation characterized by uniqueness, uncertainty, and value conflicts (Corvero, 1988) regardless of the culture in which the evaluation occurs.
REFERENCES


NOTE

This paper was originally prepared for a doctoral dissertation with the same topic by the author.
Following Up Graduates


Calculating Costs and Benefits

Training and development programs are not inexpensive. The application of training technology to enterprise personnel problems involves a substantial investment, and that investment must pay off in measurable terms. History teaches that during periods of economic downturn or recession, training budgets are among the first to be cut. Even during periods of economic growth, more and more executives are scrutinizing their training departments and holding HRD managers accountable for the cost-effectiveness of their programs.

So, like any other organizational element, training departments are being forced to justify their expenditures—and in some cases their very existence—by showing a return on investment for every dollar invested in them. Top managers speak in quantitative terms, in dollars; HRD managers, if they are to survive, must communicate their results in that language. What is needed is hard evidence that the department contributes to the achievement of enterprise profit objectives and that training and development dollars are being invested wisely. The fact is that the economic impact of well-designed and administered HRD programs is undoubtedly larger than most managers realize. Unfortunately, too many HRD departments can't supply the justification that executives demand. Cost-benefit analysis remains the exception rather than the rule in most organizations. This chapter attempts to provide a remedy for that.

Admittedly there is much to be said for the practice of treating training and development as an investment in human capital. Certainly they can be justified in terms of intangible or unquantifiable benefits—such as increased efficiency and market share or improved product quality, customer service, or competitive status. To assign such benefits a zero value would be dead wrong. But to convert them into dollar equivalents is next to impossible. At the very least, however, HRD practitioners should present unquantifiable costs and benefits along with tangible, measurable gains, tying them, when possible, to corporate strategic and operational plans, goals, and objectives.

Cost-benefit analysis is the solution. It is a technique used to compare total resources required with total benefits received from any HR program, system, service, unit, or activity. The term is also used to refer to procedures designed to provide a financial comparison between the full costs (direct, labor, and overhead) of the target
Calculating Costs and Benefits

Upon completing this chapter, the reader will be able to perform as follows:

Behavior Calculate the tangible and intangible costs and benefits of training and development programs.

Conditions Given: access to production, personnel, and financial records and reports, the results of training system evaluation, including test results, trainer and participant surveys, and on-the-job follow-up studies.

Criterion In accordance with the procedures and standards described in this chapter.

Cost Accounting Terms

The would-be training cost-benefit analyst must understand the language of cost accounting, particularly those terms that relate to training and development cost-benefit analysis. There are two categories of terms: costs and performance measures and indicators.

Costs

Capital Costs. Capital costs are expenditures for new, upgraded, rehabilitated, or new or replacement facilities or equipment for which a predetermined dollar value has been fixed and which have a relatively long life expectancy. These costs can be accounted for in two ways: purchase or lease price or cost per day or hour. In the latter case the cost is calculated by dividing the purchase or lease price by the number of days or hours the equipment or facilities will be used per year times the projected length of the lease or life of the facility or equipment in years.

Cost Centers. These are separate and distinct cost accounting entities that have been identified to provide a means of isolating and tracking costs and enhance management control by pinpointing accountability for expenditures.

Cost-per-Hire. This is a basic element of the employee recruitment budget; it is usually broken down by exempt and nonexempt jobs and often by job category. Cost-per-hire is sometimes used to measure the effectiveness of the recruitment department and individual recruiters. It includes such elements as advertising, applicant travel and subsistence, recruiter travel and subsistence, employment agency or search fees, and new employee relocation. Cost-per-hire is calculated by summing the individual element costs for each category of employee for a specified time frame (such as one year) and dividing total expenditures by the number of employees hired.

Cost-to-Complete Information. This is a category of costs often overlooked. It encompasses the costs of completing projects currently underway that may be affected by inflation, higher interest rates, or increased labor costs. It is calculated by multiplying current project expenditures in each category by forecasts of inflation, interest, and labor rate increases over the time period required to complete the project and summing the products.

Development Costs. In training, these are the costs incurred in the development
of a training system, beginning with front-end analysis and proceeding through the costs of design, development, validation, tracking, evaluation, and redesign. A sample form for the calculation of development costs is shown in Figure 22-1.

Direct Costs. Direct costs are expenses associated with specific HR projects, programs, or activities; for example, the costs incurred in conducting a workshop for supervisors. They include out-of-pocket expenses (such as air and taxi fares, subsistence, and hotel charges incurred by attending a professional conference) and direct salary costs (the salaries of instructors or consultants conducting a workshop and the salaries or wages of trainees).

Equipment Costs. The costs of purchasing (capital costs) or leasing equipment used in producing products or services. These costs are calculated by using the same formula used with capital costs:

\[
\text{Equipment costs} = \frac{\text{Purchase or lease price} \times \text{Projected length of lease}}{\text{Number of days or hours used per year} \times \text{or life of the equipment in years}}
\]

Facilities Costs. These are costs attributable to the rental of offices, classrooms, meeting rooms, shops, and laboratories.

Full Costs. Full costs are the sum of direct and indirect costs. They are the most accurate measure of how much it costs to deliver an HR service or product.

General and Administrative Costs. A category of indirect costs, such as general materials and supplies and the facilities, salaries, and benefits of administrative and support personnel.

Indirect Costs. Items of expense that are not associated with specific HR projects, programs, or activities; that is, they are costs that result from common activities that cannot be practically assigned to production as direct costs. Examples are fringe benefits (medical and health insurance, pension payments, and wellness programs) and overhead costs (such as rent, utilities, office equipment, and accounting services).

Indirect Labor. Labor needed to support the production or manufacture of a product or provide a service but not directly involved in the production or manufacture; for example, training employees in the use of equipment needed in production.

Job Cost System. This is a system for assigning production costs to products and services when production costs are of critical concern. It integrates accounting for material and labor costs, use of overhead rates, and cost accounting as a part of the formal ledger. Although time-consuming and costly because of the paperwork involved, it accurately determines factory costs of a product where a variety of products are produced.

Labor Costs. Labor costs are the portion of the total cost of a program, system, or activity that is attributable to instructors, technicians, clerical help, and trainee wages or salaries.

Materials Costs. These are the costs of the raw materials used to produce a product. Simply summing expenditures for all materials produces the figure.

 Opportunity Costs. These are the costs of forgoing other services when an HR/ HRD/OD intervention is applied in an organization. They are estimated by identifying the services that will not be performed and assigning a dollar value to those services in

(Continues on page 496.)
Calculating Costs and Benefits

Figure 22-1. A form for calculating development costs.

<table>
<thead>
<tr>
<th>1. Direct costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Administration and supervision</td>
</tr>
<tr>
<td>1. Director or coordinator</td>
</tr>
<tr>
<td>2. Instructor supervisors</td>
</tr>
<tr>
<td>3. Clerk-typists</td>
</tr>
<tr>
<td>B. Instruction</td>
</tr>
<tr>
<td>1. Instructors</td>
</tr>
<tr>
<td>2. Assistant instructors</td>
</tr>
<tr>
<td>3. Projectionists</td>
</tr>
<tr>
<td>C. Instructional materials</td>
</tr>
<tr>
<td>1. Course outlines or programs of instruction</td>
</tr>
<tr>
<td>a. Preparation or drafting</td>
</tr>
<tr>
<td>b. Typing</td>
</tr>
<tr>
<td>c. Proofing</td>
</tr>
<tr>
<td>d. Reproduction</td>
</tr>
<tr>
<td>e. Cost of materials</td>
</tr>
<tr>
<td>f. Distribution costs</td>
</tr>
<tr>
<td>2. Training aids (films, transparencies, posters, diagrams, maps, charts, models, mockups, placards)</td>
</tr>
<tr>
<td>a. Cost for purchase, and/or</td>
</tr>
<tr>
<td>b. Rental costs, and/or</td>
</tr>
<tr>
<td>c. Development costs</td>
</tr>
<tr>
<td>(1) Cost of materials</td>
</tr>
<tr>
<td>(2) Labor costs</td>
</tr>
<tr>
<td>(3) Processing costs</td>
</tr>
<tr>
<td>d. Distribution costs</td>
</tr>
<tr>
<td>3. Texts, programmed materials, handbooks, supplies, and tests</td>
</tr>
<tr>
<td>a. Commercial texts and workbooks, and programmed materials</td>
</tr>
<tr>
<td>b. Costs of contract development of texts, workbooks, programmed materials and tests</td>
</tr>
<tr>
<td>c. Costs of producing texts, workbooks, programmed materials and tests</td>
</tr>
<tr>
<td>(1) Planning</td>
</tr>
<tr>
<td>a. Job or task analyses</td>
</tr>
<tr>
<td>b. Selecting and writing training objectives</td>
</tr>
<tr>
<td>c. Constructing criterion measures</td>
</tr>
<tr>
<td>(2) Development</td>
</tr>
<tr>
<td>a. Design of materials</td>
</tr>
<tr>
<td>b. Drafting or writing tests, workbooks, program frames, and test items</td>
</tr>
<tr>
<td>c. Illustrations and art work</td>
</tr>
<tr>
<td>1. Labor</td>
</tr>
<tr>
<td>2. Materials</td>
</tr>
<tr>
<td>3. Processing</td>
</tr>
<tr>
<td>(d) Printing and typing costs</td>
</tr>
<tr>
<td>1. Labor</td>
</tr>
<tr>
<td>2. Materials</td>
</tr>
<tr>
<td>3. Processing</td>
</tr>
</tbody>
</table>

(continues)
Figure 22-1. (continued)

(e) Editing and proofing
   (f) Validating tests, programs, workbooks, tests
      1. Initial review
      2. Individual testing
         a. Writer or programmer
         b. Test population
      3. Field testing
         a. Instructor or administrator
         b. Test population
      4. Tabulation and statistical analysis
         a. ADP time
         b. Analyst time
   (g) Revision of draft materials
      1. Rewriting
      2. Review and editing
      3. Illustrations and art work
         a. Labor
         b. Materials or supplies
         c. Processing
      4. Publication costs
         a. Labor
         b. Materials
         c. Processing
      5. Retesting
         a. Administrator
         b. Test population
   (h) Tabulation and statistical analysis
      1. IH time
      2. Analyst time
   (i) Distribution costs

D. Training equipment
   1. Training aids equipment
   2. Tools and test equipment
   3. Operating gear
   4. Supplies and spare parts

II. Indirect costs
   A. Services
      1. Personnel
         a. Screening and selecting students
         b. Processing in and out
      2. Consultation in developing and reviewing curriculum materials
         a. Subject-matter experts
         b. Line supervisors
      3. Maintenance (labor costs)
         a. Setting up classrooms, shops, and laboratories, including equipment
         b. Cleaning training areas and office space
      4. Equipment, facilities repair, supplies, and spare parts
         a. Training aids equipment
         b. Operational equipment
         c. Tools and test equipment
Calculating Costs and Benefits

Figure 22-1. (continued)

<table>
<thead>
<tr>
<th>A. Facilities and equipment costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>d. Furnishings</td>
</tr>
<tr>
<td>e. Electrical and heating equipment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Utilities and facilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Electricity</td>
</tr>
<tr>
<td>2. Heat</td>
</tr>
<tr>
<td>3. Furnishings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Trainee costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Salary and benefits while in training, or</td>
</tr>
<tr>
<td>Loss of productivity in regular assignments</td>
</tr>
<tr>
<td>2. Travel and per diem</td>
</tr>
<tr>
<td>a. Transportation to training area</td>
</tr>
<tr>
<td>b. Per diem costs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. Instructor costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Recruiting costs</td>
</tr>
<tr>
<td>a. Advertising</td>
</tr>
<tr>
<td>b. Screening and selection</td>
</tr>
<tr>
<td>2. Training costs</td>
</tr>
<tr>
<td>a. Cost per instructor</td>
</tr>
<tr>
<td>b. Travel</td>
</tr>
<tr>
<td>c. Per diem</td>
</tr>
<tr>
<td>d. Supplies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E. Instructional materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reviewing and selecting commercial materials</td>
</tr>
<tr>
<td>2. Selecting contractors for development of materials</td>
</tr>
<tr>
<td>3. Consultation fees</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F. Equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cost of moving equipment</td>
</tr>
<tr>
<td>2. Estimated loss to production</td>
</tr>
</tbody>
</table>

Total cost

CALCULATION OF UNIT COSTS

<table>
<thead>
<tr>
<th>1. Cost per trainee = Total cost of system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trainees per year</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Cost per hour = Total cost of system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hours of instruction per year</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Cost per square foot = Total cost of system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of square feet of floor space</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. Equipment cost per trainee = Total cost of equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trainees in X years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. Equipment cost per hour = Total cost of equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hours of instruction in X years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. Materials cost per trainee = Total cost of materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trainees</td>
</tr>
</tbody>
</table>

* X = projected life of equipment.
terms of relinquishing profit. Opportunity costs may also be expressed in qualitative terms, such as increased customer complaints and lower employee morale.

Overhead Costs. One of the primary categories in the total costs of an HR/HRD/OD program, overhead costs consist of support staff salaries, office and classroom rentals, utilities, and the like.

Personnel Costs. Personnel costs are the "people" costs associated with an HRD program, project, or service, including both in-house personnel costs and the fees and expenses of outside consultants, trainers, and support personnel.

Travel and Subsistence Costs. These are the costs of air, bus, limo, and taxi travel, car rental and private auto, hotel or motel, meals, and incidental out-of-pocket expenses or per diem allowances.

Performance Measures and Indicators

Cost Analysis. An analytical process employed to identify and track the costs of HR activities, programs, and services, including the cost of full- and part-time company personnel, contractors, consultants, and temporary employees, real property, space and facilities, equipment, materials, supplies, and services. It encompasses the processes of collecting, validating, processing, and publishing cost histories, developing cost-estimating models from those data, conducting research to develop new cost analysis methods and techniques, and developing cost estimates during all phases of acquisition and purchasing.

Financial Utility. This is a means of evaluating an HR intervention, such as training, in financial accounting terms such as direct profit, present value, and payback period. Godlewitsch has proposed the following formulas and notes that, to be feasible, all parts of the equation must be quantified:

\[ F = N(E \times M) - C \]

where:

- \( F \) = financial utility
- \( N \) = number of people affected
- \( E \) = effect of the intervention
- \( M \) = monetary value of the effect
- \( C \) = cost of the intervention per person

To quantify the effect of an intervention, \( E \) in the formula, Godlewitsch recommends that the distribution (standard deviation) of a given skill among the subjects be measured by means of scores or ratings before and after the intervention and the shift or difference following the intervention be expressed in standard deviations, which can then be inserted into the formula.

Monetary value of the effect of the intervention \( M \) in the formula is judged by managers or calculated by framing it in terms of the standard deviation of job performance—by rule of thumb, roughly 40 percent of annual salary.

Calculating Costs and Benefits

To calculate payback (P) period in years, the formula is

\[
P = \frac{\frac{C}{E \times M}}{\frac{\text{gain}}{\text{per-person cost of training}}}
\]

To calculate return on investment, the formula is

\[
\text{ROI} = \frac{(E \times M) - C}{C} \times \text{gain} - \text{per-person cost of training}
\]

Life Cycle Cost Analysis. Analysis of the cost of a training system, organization, or item of equipment that will be incurred throughout its entire life. The cycle begins with research and development, continues through investment and operations, and extends to disposal or disestablishment of the system, organization, or equipment through redistribution, phasing out, or other disposition (scrapping or recycling).

Performance Indicators. These provide measurable evidence that a planned result has been achieved. Examples are gains in profit, productivity and reductions in interpersonal conflict, training time, costs, and the like.

Performance Measures. Performance measures are yardsticks applied to gauge whether a product or service meets performance standards and the extent to which it deviates from standard. They include assessment, inspection, personal observation, audits (quality, product, service, and procedure), questionnaires, and interviews; ratings, inventories, and checklists; diagnostic, expectancy, attitude, and morale surveys; tests and other evaluative devices; records and records; and special devices such as Gantt and PERT charts, CPM, and project charts.

Quarterly Review and Analysis (QRA). This is a control strategy used by managers to evaluate each quarter the efficiency of their organizations in the use of resources (personnel, equipment, facilities, materials, and funds) as related to progress and objectives, status, trends, deficiencies, progress, and results are reviewed, identified, analyzed, and reported to top management and subordinate departments.

Results Evaluation. This is an approach to evaluation that measures the gains or benefits of training (or other interventions) in terms of factors that affect the organization’s bottom line—such as profit or productivity.

Summative Evaluation. Summative evaluation is an approach to the evaluation of training systems that focuses on the effectiveness of the program in terms of the results obtained as compared with the resources expended.

Return on Assets (RA). An internal auditing device that indicates how well the assets of an organization are being put to use, return on assets (RA) is calculated by dividing net profit by total assets:

\[
RA = \frac{\text{net profit}}{\text{total assets}}
\]

Return on Equity (ROE). An after-tax measure used by economists to determine the quality and viability of a business, return on equity is an estimate of retained earnings for nonborrowed funds invested in a business, typically corrected for dividends paid out.
and for inflation. Or it can also be calculated by dividing net earnings by ending stockholders’ equity. Here are the two formulas:

"Real" \( RE = \frac{\text{return on equity} - \text{dividends paid out} - \text{inflation rate}}{\text{net earnings}} \) or

\[
RE = \frac{\text{net earnings}}{\text{ending stockholders' equity}}
\]

Return on Investment (ROI). In finance, ROI refers to return on invested capital and is measured as the ratio of reported income to balance sheet book value. ROI has two components: rate of turnover of total assets and rate of earnings per dollar (profit margin) of sales. It serves as a standard measure of performance for each department or element of a business. In training, ROI may be calculated as follows:

\[
ROI = \frac{\text{net program savings or benefits}}{\text{program costs}}
\]

Types of Benefits

There are two types of benefits, tangible and intangible. Tangible benefits are gains in corporate revenues or reductions in costs. Intangible benefits are gains that cannot be readily or very accurately translated into dollars or described by other concrete measures. Figure 22-3 shows in what areas tangible and intangible benefits reduce certain problems and contribute positively in certain ways.

Preliminary Requirements

In addition to the support of top management and the chief financial officer, open lines of communication between and among managers, staffers, trainers, or anyone involved in providing information and support to the cost analyst, carefully conceived plans and some critical decisions are needed prior to the initiation of training and development cost-benefit analysis. These plans and decisions address the purpose of the analysis, what will be measured, the charter of the analyst, particularly his or her authority and amount of support needed, data sources, a "charge back" accounting system, how the data will be analyzed, and how and to whom reports will go. Three items are particularly significant: the purpose of the analysis, the kind and amount of support needed, and the "charge back" accounting system.

Purpose

Cost-benefit analyses are undertaken to compare the funds expended on training and development programs with the organizational benefits derived from those expenditures. This is done to enable HRD managers and their staffs to respond to top management's requirements for hard evidence that training and development have a positive impact on the organization's bottom line. This in turn places the focus of cost-benefit analysis on
Calculating Costs and Benefits

Figure 22-2. Areas of work affected by tangible and intangible benefits.

<table>
<thead>
<tr>
<th>Tangible Benefits Decreases or Reductions in:</th>
<th>Intangible Benefits Decreases or Reductions in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absenteeism</td>
<td>Accident rates</td>
</tr>
<tr>
<td>Accidents</td>
<td>Customer complaints</td>
</tr>
<tr>
<td>Break-in time for new hires</td>
<td>Employee complaints</td>
</tr>
<tr>
<td>Grievances</td>
<td>Employee substance abuse</td>
</tr>
<tr>
<td>Machine damage</td>
<td>Labor disputes</td>
</tr>
<tr>
<td>Machine downtime</td>
<td>Production bottlenecks</td>
</tr>
<tr>
<td>Labor costs</td>
<td>Violations of policy</td>
</tr>
<tr>
<td>Operating costs</td>
<td>Work backlogs</td>
</tr>
<tr>
<td>Rejects and reworks</td>
<td></td>
</tr>
<tr>
<td>Tardiness</td>
<td></td>
</tr>
<tr>
<td>Turnover</td>
<td></td>
</tr>
<tr>
<td>Unit costs</td>
<td></td>
</tr>
<tr>
<td>Waste and spoilage</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Increases or Improvements in:</th>
<th>Increases or Improvements in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance</td>
<td>Customer relations</td>
</tr>
<tr>
<td>Cost benefits</td>
<td>Customer satisfaction</td>
</tr>
<tr>
<td>Production</td>
<td>Customer services</td>
</tr>
<tr>
<td>Productive time</td>
<td>Employee job satisfaction</td>
</tr>
<tr>
<td>Profit</td>
<td>Employee motivation and morale</td>
</tr>
<tr>
<td>Return on assets</td>
<td>Employee skills</td>
</tr>
<tr>
<td>Return on equity</td>
<td>Employee suggestions</td>
</tr>
<tr>
<td>Return on investment</td>
<td>New processes and products</td>
</tr>
<tr>
<td>Sales</td>
<td>New production methods</td>
</tr>
<tr>
<td></td>
<td>Pool of promotables</td>
</tr>
<tr>
<td></td>
<td>Productivity</td>
</tr>
<tr>
<td></td>
<td>Product quality</td>
</tr>
<tr>
<td></td>
<td>Responsiveness to new needs</td>
</tr>
</tbody>
</table>

Tangible benefits—gains in corporate revenues or reductions in costs. This does not preclude identifying and reporting intangible benefits—gains that cannot be described by dollars or other concrete measures. Such benefits, however, should only be used to supplement the tangible evidence.

Support Needed

To be timely and effective, adequate resources must be provided to the analyst. Such support includes people (whether assigned to the HRD department or another line or staff element), time (for the analyst, staffers, and others to conduct and report the analysis), funding (to develop the system, required instruments, and so on), equipment (computers, scanners, and the like), and access (to people, reports, and records). Without support, realistic and accurate cost-benefit analysis cannot be performed.
Accounting System

More and more HRD departments are being converted from staff entities to cost centers. That means that the department must, at a minimum, cover its own costs (such as staff salaries, vendors, overhead, supplies, and equipment) and, ideally, show a profit. In most organizations, the creation of an HRD cost center precipitates a charge-back accounting system, whereby the users of training and development in other departments and staff sections of the company are charged a fee for HRD programs and services. The fee is usually competitive with the fees charged by outside vendors and consultants.

Although most corporate users of training and other HRD services would prefer a no-fee system, a "pay for services" system has one big advantage—a positive and lasting effect on the corporate image of training. More than that, it saves corporate funds by forcing managers to make responsible budget allocation decisions by considering the real need for training. Charge-back systems also practically eliminate training "no shows" and class cancellations due to low enrollment in scheduled training courses.

One caveat: Be sure that "revenue" from your charge-back system doesn't become more important than corporate requirements for timely quality training for individuals and groups.

Sources of Data

What are the sources of costs and benefits information? How do you collect the data? Essentially there are three sources: existing corporate records and reports, HR department records, and directed accounting surveys and studies.

Corporate Records and Reports

All organizations produce, distribute, and retain in files a great variety of production, financial, marketing and sales, and personnel reports and records. Others deal with labor relations, employee relations, grievances, suggestions, and special studies. Some records are held in corporate headquarters' staff offices; others are maintained in branch offices and production facilities. Many records are stored in computer files so that retrieval is not a problem.

HR Department Records

Since the mid-nineteen thirties, federal and state laws regulating wages and conditions of employment have been enacted. This legislation has created a need for accurate record-keeping systems with quick retrieval capability for reporting purposes. In addition to having records available to meet government reporting requirements, most HR organizations maintain an efficient human resources information system (HRIS) so that all types of services can be provided to employees and information can be made available (on a need-to-know basis) to managers and staff officers. These records can be accessed to provide specific information for cost-benefit analyses.

In addition, any forward-looking HR department maintains records of employee performance in training and development programs as well as the results of follow-up when the employee returns to the job. These records will prove to be very useful.
Calculating Costs and Benefits

Figure 22.4. A simple method of calculating cost per applied person-day.

<table>
<thead>
<tr>
<th>Step</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Calculate labor cost/day:</td>
<td>Annual salary ( \div ) No. days worked/yr = Labor cost/day ( \left( \frac{\text{dollars}}{\text{day}} \right) )</td>
</tr>
<tr>
<td>2. Calculate full cost/day:</td>
<td>Labor cost/day ( \times ) OHD factor = Full cost/day ( \left( \frac{\text{dollars}}{\text{day}} \right) )</td>
</tr>
<tr>
<td>3. Calculate applied rate:</td>
<td>No. days/planned on project ( \div ) No. working days/no. project = Applied rate ( \left( \frac{\text{dollars}}{\text{day}} \right) )</td>
</tr>
<tr>
<td>4. Calculate cost per applied person-day:</td>
<td>Full cost/day ( \div ) Applied rate ( \left( \frac{\text{dollars}}{\text{day}} \right) ) = Cost applied person-day ( \left( \frac{\text{dollars}}{\text{day}} \right) )</td>
</tr>
</tbody>
</table>


Vented to dollars, are tallied following the training. The full costs of the training are then subtracted from the increased revenue to yield the net benefit.

**Equipment.** Potential cost benefits in the equipment area are achievable by training and development interventions and can be realized in terms of (1) increases in production and (2) avoidance of the costs associated with equipment downtime.

To calculate benefits in terms of both increases in productivity and reductions in equipment downtime, the first step is to identify the worth or value of the equipment. To arrive at the productivity increase, determine the number of hours, days, or weeks of usage prior to the intervention and following the intervention, find the difference between them, and multiply the result by the value of the equipment. Finally, subtract the full costs of the training to arrive at the final cost-benefit calculation.

To determine the costs avoided by equipment downtime, again calculate the worth or value of the equipment, determine the number of hours or days of downtime prior to and following the intervention, multiply the difference by the value of the equipment, and add the difference in the costs of repairing the equipment prior to and after the training. Then subtract the full costs of the training to arrive at the cost-benefit figure.

**Equipment Repair and Maintenance.** To show reductions in equipment repair and maintenance costs attributable to training (operator preventive maintenance or repairer training), determine the full costs of repair and maintenance over a specified period of time (usually one year) provided by in-house personnel, outside vendors, or both. Then add the value of reduced equipment downtime over the same period, and subtract the full costs of the training.

**Travel and Transportation Management.** Training of managers and supervisors in travel and transportation management policies and procedures can pay high dividends.
Figure 22-5. Calculating the cost of an HR service, program, or project.

<table>
<thead>
<tr>
<th>LABOR COSTS</th>
<th>Person-days</th>
<th>Cost/Person-day*</th>
<th>Subtotals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HR Employees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Professional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Clerical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Administrative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Clients and participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Manager</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Exempt participants (average salary)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Nonexempt participants (average salary)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Labor Cost</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIRECT COSTS</th>
<th>No. of Days or Units</th>
<th>Cost/Day or Unit</th>
<th>Subtotals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Travel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. HR personnel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Air</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Ground</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Clients and participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Air</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Ground</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Per diem expenses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. HR personnel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Clients and participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Materials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Purchased services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Equipment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Facilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Other costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Direct Cost</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL COST/FPHASE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Cost/day is used most often, but cost/hour, week, year, or term of supply (such as cost/workbooks) should be substituted as appropriate.

Calculating Costs and Benefits

Cost savings attributable to that training can be easily calculated by determining the full corporate or branch costs of travel and transportation for a specified time frame prior to the intervention and following the intervention and then subtracting the sum of the full costs of the training and posttraining travel and transportation costs for a comparable time period from the pretraining costs.

**Personnel Replacement (Turnover) Costs.** Cascio has identified two methods of human resource accounting: asset models and expense or utility models.³

The asset model highlights all costs associated with corporate human resources. It typically employs one of two methodologies: historical cost and replacement cost approaches. Both involve expert (usually manager) estimates of the costs associated with specified employee-related actions. For example, with the historical cost method, managers estimate the costs of recruitment, selection, training, and other HR programs, services, and activities for a given time period. The resulting figures are summed to establish the costs associated with human resources for that organization.

The replacement cost method requires managers to identify costs associated with the loss and replacement of an employee, such as lost revenue due to substandard performance, remedial training for an incumbent who failed, training for a new incumbent, downtime between incumbent changeover, start-up time for new hire, relocation expenses for new employees, and staff and legal costs of justifying termination.⁴ To calculate the replacement cost, sum the costs of the foregoing factors for one individual in a particular worker category and multiply the result by the number of workers in that category (for example, electronics technicians) who were prevented from failing by providing training. (See Figure 22-6.)

The expense model measures the economic consequences to an organization of a worker's behavior by determining the utility of training programs. Utility lies in the ability of a training program to improve the effectiveness of participants beyond their pretraining level of competence.

According to Sheppeck and Cohen, several assumptions must be made before determining the utility of a training program:

- Differences in employee performance occur on most jobs.
- Training programs produce improved employee performance.
- Increases in employee performance result in increased company profitability.

To calculate utility, a function of the duration of a training program's effect on employees, the following factors are considered: number of workers trained, validity of the training program, the value of the job for which training was provided, and the total cost of the program (see Figure 23-7). Two of those factors, validity and value, are determined as described below.

**Validity of a Training Program.** In cost-benefit analysis, validity is the difference in performance between trained and untrained workers as determined by supervisors' ratings of the performance of each group, quantitative and qualitative production and sales measures, or service indicators. Obviously, the latter two are more objective.

Figure 22-6. Example of the calculation of training program economic benefit as determined by the replacement cost method.

\[
\text{Total Savings to the Organization} = \text{No. of Supervisors Prevented From Failing} \times \text{Replacement Cost per Supervisor} \times \text{No. of Supervisors Trained} \times \text{Training Cost per Supervisor} - \text{Total Cost of Training}
\]

\[
\begin{array}{c}
\text{5} \\
\text{\$57,000} \\
\text{100} \\
\text{\$500}
\end{array}
\]

\[
= \text{\$285,000} \text{ Minus} \text{\$50,000}
\]

Total Economic Benefit to the Organization: \text{\$235,000}

From Michael A. Shepperd and Stephen L. Cohen, "Put a Dollar Value on Your Training Programs," Training and Development Journal, November 1985, p. 60. Copyright 1985 the American Society for Training and Development. Reprinted with permission. All rights reserved.

Figure 22-7. A formula for calculating utility.

\[
\text{Utility} = \left( \frac{\text{Years Duration} \times \text{Effect on Performance}}{\text{Trained}} \right) \times \left( \frac{\text{Performance Difference}}{\text{Value} - \text{Trained} \times \text{Untrained Employees}} \right) \times \left( \frac{\text{Number of Employees}}{\text{Cost per Employee}} \right)
\]

From Michael A. Shepperd and Stephen L. Cohen, "Put a Dollar Value on Your Training Programs," Training and Development Journal, November 1985, p. 60. Copyright 1985 the American Society for Training and Development. Reprinted with permission. All rights reserved.

measures but are not always appropriate measures of job performance. Therefore, supervisory performance ratings should be obtained before and after training and compared with a control group of employees who have not had the training. (See Figure 22-8 for an example of a utility calculation.)

Value of the Job. In cost-benefit analysis, job "value" is the difference in the dollar contribution to the company of high versus low employee performance, sometimes referred to as the standard deviation of job performance in dollars. It represents the value to the company of improving an employee's job performance from average to
Calculating Costs and Benefits

high. Value is influenced by the extent to which the job affects the quality and quantity of a product and the nature of the product. For example, HR jobs usually affect product quality more than clerical-administrative jobs.

Value is calculated by having supervisors or other subject-matter experts estimate the yearly value to the organization of the products and services for outstanding, average, and marginal employees. The costs of having an external source provide the products and services is used as a benchmark by the judges. The average of the estimates of all experts is obtained for each level of performance, and the difference between the outstanding and average or the average and marginal performance is considered "value."]

Other People-Problem Interventions. Another potentially lucrative area for cost-benefit analyses is the calculation of benefits accruing to the solution of such people problems as absenteeism, accidents, injuries, grievances, interpersonal conflicts, strikes, and tardiness. The calculation is relatively straightforward once the full costs of the people involved have been determined. Steps in the process follow:

1. Select the problem or event.
2. Identify the employees involved in the problem.
3. Calculate the full cost of each employee in terms of hours, days, or weeks.
4. Compute total people costs by multiplying the number of people in each category involved by their cost per day, hour, or week and summing the products.
5. Subtract the full costs of the HRD intervention from the total costs of the problem or occurrence.

Computer-Based Software

Fortunately, not all of the computations needed to calculate training and development cost-benefits need to be done manually. Several computer-based tools, including spread-

Figure 22-8. Calculation of the utility of a supervisory training program in participative management.

\[
\text{Utility} = \frac{\text{(Years Duration of Effect on Performance) \times (Number Trained) \times (Performance Difference Between Trained and Untrained Employees) \times (Value) - (Cost)}}{20 \times \frac{75}{20} \times \frac{15,000}{20} \times \frac{1,000}{20}} = \frac{400,000}{20} = 400,000
\]

sheets, are now commercially available to help with the tasks of collecting and analyzing data and performing different calculations. And other word processing applications can help with the preparation of reports and visuals for information and decision briefings. Figure 22-9 shows Kearsley's "Models in Cost/Benefit Disk." Figures 22-10 and 22-11 show breakdowns of the data required for Kearsley’s CBT Cost Justification and Training Cost models. (Sources of other cost-benefit software are listed in Appendix B.)

Reporting

Cost-benefit analyses do little to justify HRD expenditures unless they are reported to line and staff managers, top executives, and the board of control. That is best accomplished by several complementary means: formal written reports, formal information briefings (after the fact), formal decision briefings (before the intervention), quarterly review and analysis reports, and informal memorandums, status reports, briefings, discussions, and meetings.

Although the amount of detail included in presenting the results of cost-benefit

Figure 22-9. Models in cost/benefits disk.

| Resource Requirements—Compare different training approaches in terms of resources required across different training activities. |
| Life Cycle—Calculate the costs of different components of a training program during each phase of a product or system. |
| Return on Investment—Compute the ratio of training costs to the estimated or measured value of training. |
| Break Even—Compare the break-even point between two different training approaches in terms of costs and outcomes. |
| Compensation—Calculate the annual costs of employee compensation due to poor health or accidents. |
| Productivity—Plot the productivity functions relating training costs and outcomes for up to three different approaches. |
| Unit Cost—Compute training costs on the basis of each student or each student-hour. |
| Resource Estimation—Estimate the optimum number of training devices needed for a given training program. |
| Transfer—Calculate the cost savings due to the use of simulation instead of actual equipment for training. |
| Benefits—Analyze the causal relationships between a set of training activities and training outcomes. |

Figure 22-10. Data required in the CBT cost justification model.

<table>
<thead>
<tr>
<th>Classroom Training Course Development Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course length</td>
</tr>
<tr>
<td>Course development time</td>
</tr>
<tr>
<td>Number of students requiring training</td>
</tr>
<tr>
<td>Maximum number of students</td>
</tr>
<tr>
<td>Cost of student handouts</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classroom Training Labor Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum developer rate</td>
</tr>
<tr>
<td>Instructor rate</td>
</tr>
<tr>
<td>Student rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classroom Training Travel Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student travel</td>
</tr>
<tr>
<td>Instructor travel</td>
</tr>
<tr>
<td>Percentage of students travelling</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBT Course Development Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course length</td>
</tr>
<tr>
<td>Course development time</td>
</tr>
<tr>
<td>Authoring system cost</td>
</tr>
<tr>
<td>Authoring system lifetime</td>
</tr>
<tr>
<td>Authoring system usage</td>
</tr>
<tr>
<td>CBT developer rate</td>
</tr>
</tbody>
</table>


Analyses will depend on the scope and complexity of the analyses, most reports should include the following:

- Definition of the purpose of the analysis
- Description of specific elements of the HRD program or service analyzed
- Identification of the data collection and reduction techniques used
- Description of evaluation and analysis activities
- Identification of findings and conclusions reached
- Delineation of recommendations for action

**Procedures**

1. Step 1. Define the specific purposes and objectives of the analysis.
2. Step 2. Develop an analysis plan.
   - a. What will be measured
Figure 22-11. Data required in the training cost model.

- Number of people to be trained in group
- Number of actual working days in each year
- Average annual cost (of salaries plus benefits)
- Length of training session (in days)
- Related travel time (in days)
- Average cost of travel
- Average cost of accommodation
- Annual cost of training (salary + overhead)
- Classroom course development cost
- Student package cost (replication)
- Classroom trainer/student ratio
- Working days in each year for trainer
- Interactive video course development cost
- Student/workstation ratio
- Annual workstation lease charge
- Session length reduction (percent)
- Number of times course is run


b. Sources of data
c. Accounting system and computer-based tools to be used
d. How data will be analyzed
e. When and how results will be reported and to whom

Step 5. Communicate the purpose, objectives, and plans to all constituencies.

Step 4. Get support and resources for the project.
a. Top management and the chief financial officer
b. Line managers and heads of staff
c. HRD staffs
d. Authority and access for analyst(s)
e. Time
f. Funding
g. Equipment

Step 5. Collect, and reduce the data.
a. Current costs of the system, program, or service
b. Costs of the HRD intervention (actual or projected)
c. Benefits accruing to the HRD intervention (actual or projected)

Step 6. Analyze the data.
a. Compare costs and benefits.
b. Develop conclusions and recommendations.

Step 7. Report findings to all constituencies.
a. Prepare a formal written report.
b. Prepare and present an information briefing (after the fact) or a decision briefing (before the intervention).

Step 8. Carry out recommendations or decisions reached.
CHECKLIST FOR CALCULATING COSTS AND BENEFITS

- Is adequate support provided for cost-benefit analysis in terms of:
  - Top management sponsorship?
  - Backing by the chief financial officer?
  - Open lines of communication and access to people, reports, and records?
  - Personnel?
  - Time?
  - Funding?
  - Equipment?

- Are the following sources of costs and benefits data exploited:
  - Corporate records and reports?
  - HR Department records and reports, including HRIS?
  - Special studies?

- Have all training and development costs been identified, tracked, and recorded:
  - Capital costs?
  - Costs-per-hire?
  - Costs-to-complete?
  - Development costs?
  - Direct costs?
  - Equipment costs?
  - Facilities costs?
  - Full costs?
  - General and administrative costs?
  - Indirect costs?
  - Indirect labor?
  - Labor costs?
  - Materials costs?
  - Opportunity costs?
  - Overhead costs?
  - Personnel costs?
  - Travel and subsistence costs?

- Have unit costs of the following types for training been calculated:
  - Cost per trainee?
  - Cost per hour?
  - Cost per square foot?
  - Equipment cost per trainee?
  - Equipment cost per hour?
  - Materials cost per trainee?

- Have the following types of performance measures and indicators been identified, tracked, and recorded:
  - Gains in profit, product quality, and productivity?
  - Return on assets, equity, and investment?
  - Reductions in absenteeism, accidents, machine damage, training costs, training time, and turnover?
Designing Training and Development Systems

- Decreases in accident rates, labor disputes, production bottlenecks, and work backlogs?
- Improvements in customer satisfaction, customer services, employee job satisfaction, suggestions, and product quality?
- Results of the administration of tests, surveys, and other evaluative devices?

- Are the results of cost-benefit analysis reported to all of the following:
  - The board of control?
  - Top management?
  - Line and staff managers?

- Do reports contain the following items:
  - The purpose and coverage of the analysis?
  - Data collection, reduction, and analysis methodology used?
  - Findings and conclusions?
  - Recommendations for action?

For Further Reading, Viewing, and Listening


Calculating Costs and Benefits


Chapter 3

The Four Levels:
An Overview

The four levels represent a sequence of ways to evaluate programs. Each level is important and has an impact on the next level. As you move from one level to the next, the process becomes more difficult and time-consuming, but it also provides more valuable information. None of the levels should be bypassed simply to get to the level that the trainer considers the most important. These are the four levels:

Level 1—Reaction
Level 2—Learning
Level 3—Behavior
Level 4—Results

Reaction

As the word reaction implies, evaluation on this level measures how those who participate in the program react to it. I call it a measure of customer satisfaction. For many years, I conducted seminars, institutes, and conferences at the University of Wisconsin Management Institute. Organizations paid a fee to send their people to these public programs. It is obvious that the reaction of participants was a measure of customer satisfaction. It is also obvious that reaction had to be favorable if we were to stay in business and attract new customers as well as get present customers to return to future programs.
Evaluating Training Programs

It isn’t quite so obvious that reaction to in-house programs is also a measure of customer satisfaction. In many in-house programs, participants are required to attend whether they want to or not. However, they still are customers even if they don’t pay, and their reactions can make or break a training program. What they say to their bosses often gets to higher-level managers, who make decisions about the future of training programs. So, positive reactions are just as important for trainers who run in-house programs as they are for those who offer public programs.

It is important not only to get a reaction but to get a positive reaction. As just described, the future of a program depends on positive reaction. In addition, if participants do not react favorably, they probably will not be motivated to learn. Positive reaction may not ensure learning, but negative reaction almost certainly reduces the possibility of its occurring.

Learning

Learning can be defined as the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending the program.

Those are the three things that a training program can accomplish. Programs dealing with topics like diversity in the workforce aim primarily at changing attitudes. Technical programs aim at improving skills. Programs on topics like leadership, motivation, and communication can aim at all three objectives. In order to evaluate learning, the specific objectives must be determined.

Some trainers say that no learning has taken place unless change in behavior occurs. In the four levels described in this book, learning has taken place when one or more of the following occurs: Attitudes are changed. Knowledge is increased. Skill is improved. One or more of these changes must take place if a change in behavior is to occur.

Behavior

Behavior can be defined as the extent to which change in behavior has occurred because the participant attended the training program. Some trainers want to bypass levels 1 and 2—reaction and learning—in order to measure changes in behavior. This is a serious mistake. For example,
The Four Levels: An Overview

suppose that no change in behavior is discovered. The obvious conclusion is that the program was ineffective and that it should be discontinued. This conclusion may or may not be accurate. Reaction may have been favorable, and the learning objectives may have been accomplished, but the level 3 or 4 conditions may not have been present.

In order for change to occur, four conditions are necessary:

1. The person must have a desire to change.
2. The person must know what to do and how to do it.
3. The person must work in the right climate.
4. The person must be rewarded for changing.

The training program can accomplish the first two requirements by creating a positive attitude toward the desired change and by teaching the necessary knowledge and skills. The third condition, right climate, refers to the participant's immediate supervisor. Five different kinds of climate can be described:

1. Preventing: The boss forbids the participant from doing what he or she has been taught to do in the training program. The boss may be influenced by the organizational culture established by top management. Or the boss's leadership style may conflict with what was taught.

2. Discouraging: The boss doesn't say, "You can't do it," but he or she makes it clear that the participant should not change behavior because it would make the boss unhappy. Or the boss doesn't model the behavior taught in the program, and this negative example discourages the subordinate from changing.

3. Neutral: The boss ignores the fact that the participant has attended a training program. It is business as usual. If the subordinate wants to change, the boss has no objection as long as the job gets done. If negative results occur because behavior has changed, then the boss may turn into a discouraging or even preventing climate.

4. Encouraging: The boss encourages the participant to learn and apply his or her learning on the job. Ideally, the boss discussed the program with the subordinate beforehand and stated that the two would discuss application as soon as the program was over. The boss basically says, "I am interested in knowing what you learned and how I can help you transfer the learning to the job."

5. Requiring: The boss knows what the subordinate learns and makes sure that the learning transfers to the job. In some cases, a learning
contract is prepared that states what the subordinate agrees to do. This contract can be prepared at the end of the training session, and a copy can be given to the boss. The boss sees to it that the contract is implemented. Malcolm Knowles’s book *Using Learning Contracts* (San Francisco: Jossey-Bass, 1986) describes this process.

The fourth condition, rewards, can be intrinsic (from within), extrinsic (from without), or both. Intrinsic rewards include the feelings of satisfaction, pride, and achievement that can occur when change in behavior has positive results. Extrinsic rewards include praise from the boss, recognition by others, and monetary rewards, such as merit pay increases and bonuses.

It becomes obvious that there is little or no chance that training will transfer to job behavior if the climate is preventing or discouraging. If the climate is neutral, change in behavior will depend on the other three conditions just described. If the climate is encouraging or requiring, then the amount of change that occurs depends on the first and second conditions.

As stated earlier, it is important to evaluate both reaction and learning in case no change in behavior occurs. Then it can be determined whether the fact that there was no change was the result of an ineffective training program or of the wrong job climate and lack of rewards.

It is important for trainers to know the type of climate that participants will face when they return from the training program. It is also important for them to do everything that they can to see to it that the climate is neutral or better. Otherwise there is little or no chance that the program will accomplish the behavior and results it seeks, because participants will not even try to use what they have learned. Not only will no change occur, but those who attended the program will be frustrated with the boss, the training program, or both for teaching them things that they can’t apply.

One way to create a positive job climate is to involve bosses in the development of the program. Chapter 1 suggested asking bosses to help to determine the needs of subordinates. Such involvement helps to ensure that a program teaches practical concepts, principles, and techniques. Another approach is to present the training program, or at least a condensed version of it, to the bosses before the supervisors are trained.

A number of years ago, I was asked by Dave Harris, personnel manager, to present an eighteen-hour training program to 240 supervisors at A. O. Smith Corporation in Milwaukee. I asked Dave if he could
arrange for me to present a condensed, three- to six-hour version to the company’s top management. He arranged for the condensed version to be offered at the Milwaukee Athletic Club. After the six-hour program, the eight upper-level managers were asked for their opinions and suggestions. They not only liked the program but told us to present the entire program first to the thirty-five general foremen and superintendents who were the bosses of the 240 supervisors. We did what they suggested. We asked these bosses for their comments and encouraged them to provide an encouraging climate when the supervisors had completed the program. I am not sure to what extent this increased change in behavior over the level that we would have seen if top managers had not attended or even known the content of the program, but I am confident that it made a big difference. We told the supervisors that their bosses had already attended the program. This increased their motivation to learn and their desire to apply their learning on the job.

Much has been written concerning change in behavior, or “transfer of training,” as it is often termed. Some of the references at the end of Chapter 8 describe concepts, principles, and techniques.

Results

Results can be defined as the final results that occurred because the participants attended the program. The final results can include increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, increased sales, reduced turnover, and higher profits. It is important to recognise that results like these are the reason for having some training programs. Therefore, the final objectives of the training program need to be stated in these terms.

Some programs have these in mind on a long-term basis. For example, one major objective of the popular program on diversity in the workforce is to change the attitudes of supervisors and managers toward minorities in their departments. We want supervisors to treat all people fairly, show no discrimination, and so on. These are not tangible results that can be measured in terms of dollars and cents. But it is hoped that tangible results will follow. Likewise, it is difficult if not impossible to measure final results for programs on such topics as leadership, communication, motivation, time management, empowerment,
decision making, or managing change. We can state and evaluate desired behaviors, but the final results have to be measured in terms of improved morale or other nonfinancial terms. It is hoped that such things as higher morale or improved quality of work life will result in the tangible results just described.

Summary

Trainers should begin to plan by considering the desired results. These results should be determined in cooperation with managers at various levels. Surveys and/or interviews can be used. A desirable and practical approach is to use an advisory committee consisting of managers from different departments. Their participation will give them a feeling of ownership and will probably increase the chances of their creating a climate that encourages change in behavior. The next step is to determine what behaviors will produce the desired results. Then trainers need to determine what knowledge, skills, and attitudes will produce the desired behavior.

The final challenge is to present the training program in a way that enables the participants not only to learn what they need to know but also to react favorably to the program. This is the sequence in which programs should be planned. The four levels of evaluation are considered in reverse. First, we evaluate reaction. Then, we evaluate learning, behavior, and results—in that order. Each of the four levels is important, and we should not bypass the first two in order to get to levels 3 and 4. Reaction is easy to do, and we should measure it for every program. Trainers should proceed to the other three levels as staff, time, and money are available. The next four chapters provide guidelines, suggested forms, and procedures for each level. The case studies in Part Two of the book describe how the levels were applied in different types of programs and organizations.
Chapter 4

Evaluating Reaction

Evaluating reaction is the same thing as measuring customer satisfaction. If training is going to be effective, it is important that trainees react favorably to it. Otherwise, they will not be motivated to learn. Also, they will tell others of their reactions, and decisions to reduce or eliminate the program may be based on what they say. Some trainers call the forms that are used for the evaluation of reaction happiness sheets. Although they say this in a critical or even cynical way, they are correct. These forms really are happiness sheets. But they are not worthless. They help us to determine how effective the program is and learn how it can be improved.

Measuring reaction is important for several reasons. First, it gives us valuable feedback that helps us to evaluate the program as well as comments and suggestions for improving future programs. Second, it tells trainees that the trainers care about helping them do their job better and that they need feedback to determine how effective they are. If we do not ask for reaction, we told to the trainers that we know what they want and need and that we can judge the effectiveness of the program without getting feedback from them. Third, reaction sheets can provide quantitative information that you can give to managers and others concerned about the program. Finally, reaction sheets can provide trainers with quantitative information that can be used to establish standards of performance for future programs.

Evaluating reaction is not only important but also easy to do and do effectively. Most trainers use reaction sheets. I have seen dozens of forms and various ways of using them. Some are effective, and some
are not. Here are some guidelines that will help trainers to get maximum benefit from reaction sheets:

Guidelines for Evaluating Reaction
1. Determine what you want to find out.
2. Design a form that will quantify reactions.
3. Encourage written comments and suggestions.
4. Get 100 percent immediate response.
5. Get honest responses.
6. Develop acceptable standards.
7. Measure reactions against standards, and take appropriate action.
8. Communicate reactions as appropriate.

The next eight sections contain suggestions for implementing each of these guidelines.

Determine What You Want to Find Out

In every program, it is imperative to get reactions both to the subject and to the leader. And it is important to separate these two ingredients of every program. In addition, trainers may want to get trainees' reactions to one or more of the following: the facilities (location, comfort, convenience, and so forth); the schedule (time, length of program, breaks, convenience, and so forth); meals (amount and quality of food and so forth); case studies, exercises, and so forth; audiovisual aids (how appropriate, effective, and so forth); handouts (how helpful, amount, and so forth); the value that participants place on individual aspects of the program.

Design a Form That Will Quantify Reactions

Trainers have their own philosophy about the forms that should be used. Some like open questions that require a lot of writing. They feel that checking boxes does not provide enough feedback. Some even feel that it amounts to telling trainees what to do. Others keep it as simple as possible and just ask trainees to check a few boxes.

The ideal form provides the maximum amount of information and requires the minimum amount of time. When a program is over, most trainees are anxious to leave, and they don't want to spend a lot of time
Evaluating Reaction 27

completing evaluation forms. Some even feel that trainers do not consider their comments anyway.

There are a number of different forms that can provide the maximum information and require a minimum amount of time to complete. Exhibits 4.1, 4.2, 4.3, and 4.4 show forms that can be used

Exhibit 4.1. Reaction Sheet

Please give us your frank reactions and comments. They will help us to evaluate this program and improve future programs.

Leader ____________________________ Subject ____________________________

1. How do you rate the subject? (interest, benefit, etc.)
   _____ Excellent
   _____ Very good
   _____ Good
   _____ Fair
   _____ Poor

   Comments and suggestions:

2. How do you rate the conference leader? (knowledge of subject matter, ability to communicate, etc.)
   _____Excellent
   _____ Very good
   _____ Good
   _____ Fair
   _____ Poor

   Comments and suggestions:

3. How do you rate the facilities? (comfort, convenience, etc.)
   _____ Excellent
   _____ Very good
   _____ Good
   _____ Fair
   _____ Poor

   Comments and suggestions:

4. How do you rate the schedule?
   _____ Excellent
   _____ Very good
   _____ Good
   _____ Fair
   _____ Poor

   Comments and suggestions:

5. What would have improved the program?

   ________________________________________________________________

   ________________________________________________________________
28  Evaluating Training Programs

Exhibit 4.2. Reaction Sheet

Leader __________________________ Subject __________________________

1. How pertinent was the subject to your needs and interests?
   _____ Not at all _______ To some extent _______ Very much

2. How was the ratio of presentation to discussion?
   _____ Too much presentation _______ Okay _______ Too much discussion

3. How do you rate the instructor?
   a. In stating objectives Excellent | Very good | Good | Fair | Poor |
   b. In keeping the session alive and interesting
   c. In communicating
   d. In using aids
   e. In maintaining a friendly and helpful attitude

4. What is your overall rating of the leader?
   _____ Excellent  Comments and suggestions:
   _____ Very good
   _____ Good
   _____ Fair
   _____ Poor

5. What would have made the session more effective?

________________________

effectively when one leader conducts the entire program. Exhibit 4.5 is unusual because it is truly a "smile" sheet, as many reaction sheets are called. I found it in a hotel in Geneva, Switzerland. The original form was written in French. Exhibits 4.5 and 4.6 show forms that can be used when more than one leader conducts the program and it is not desirable to have trainees complete a separate form for each. All forms
Evaluating Reaction

Exhibit 4.3. Reaction Sheet

In order to determine the effectiveness of the program in meeting your needs and interests, we need your input. Please give us your reactions, and make any comments or suggestions that will help us to serve you.

*Instructions:* Please circle the appropriate response after each statement.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The material covered in the program was relevant to my job.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The material was presented in an interesting way.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The instructor was an effective communicator.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The instructor was well prepared.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The audiovisual aids were effective.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. The handouts will be of help to me.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I will be able to apply much of the material to my job.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. The facilities were suitable.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. The schedule was suitable.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. There was a good balance between presentation and group involvement.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I feel that the workshop will help me do my job better.</td>
<td>1 2 3 4 5 6 7 8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What would have improved the program?

---

can be quantified and used to establish standards for future evaluations. It would be worthwhile to try a form with several groups to see whether trainees understand it and whether it serves the purpose for which it was designed. All the forms illustrated in this chapter need to be tabulated by hand. They can be readily adapted so that they can be tabulated and analyzed by computer if that is easier.
### Evaluating Training Programs

#### Exhibit 4.4. Reaction Sheet

Please complete this form to let us know your reaction to the program. Your input will help us to evaluate our efforts, and your comments and suggestions will help us to plan future programs that meet your needs and interests.

**Instructions:** Please circle the appropriate number after each statement and then add your comments.

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How do you rate the subject content? (interesting, helpful, etc.)</td>
<td>5 4 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>2. How do you rate the instructor? (preparation, communication, etc.)</td>
<td>5 4 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>3. How do you rate the facilities? (comfort, convenience, etc.)</td>
<td>5 4 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>4. How do you rate the schedule? (time, length, etc.)</td>
<td>5 4 3 2 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments:</td>
<td></td>
</tr>
<tr>
<td>5. How would you rate the program as an educational experience to help you do your job better?</td>
<td>5 4 3 2 1</td>
<td></td>
</tr>
<tr>
<td>6. What topics were most beneficial?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. What would have improved the program?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dear Client,

We would like to have your comments and suggestions to enable us to offer you the kind of service you would like.

Would you help us by ticking the face that is most indicative of your feelings:

☐ breakfast ☐ lunch  Very good Good Average

1. Are you satisfied with the quality of the meals?

2. Are you satisfied with the variety of dishes available?

3. Do you find our prices competitive?

4. What do you think of the service?

5. How do you find the atmosphere in the restaurant?

6. Suggestions:

Name: ____________________________

Address: ____________________________
Exhibit 4.6. Reaction Sheet

Please give your frank and honest reactions. Insert the appropriate number.

Scale: 5 = Excellent 4 = Very good 3 = Good 2 = Fair 1 = Poor

<table>
<thead>
<tr>
<th>Leader</th>
<th>Subject</th>
<th>Presentation</th>
<th>Discussion</th>
<th>Audiovisual aids</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom Jones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeraldFord</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luis Aparicio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simon Bolivar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muhammad Ali</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chris Columbus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bert Starr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Facilities Rating _____ Meals Rating _____
Comments: Comments:

Schedule Rating _____ Overall program Rating _____
Comments: Comments:

What would have improved the program?

Encourage Written Comments and Suggestions

The ratings that you tabulate provide only part of the participants' reactions. They do not provide the reasons for those reactions or suggest what can be done to improve the program. Therefore, it is important to get additional comments. All the forms shown in this chapter give participants opportunities to comment.

Typically, reaction sheets are passed out at the end of a program. Participants are encouraged to complete the forms and leave them on the back table on their way out. If they are anxious to leave, most will not take time to write in their comments. You can prevent this by
making the completion of reaction sheets part of the program. For example, five minutes before the program is scheduled to end, the instructor can say: “Please take time to complete the reaction sheet, including your comments. Then I have a final announcement.” This simple approach will ensure that you receive comments from all or nearly all the participants.

Another approach is to pass the forms out at the beginning of the program and stress the importance of comments and suggestions.

Get 100 Percent Immediate Response

I have attended many programs at which reaction sheets are distributed to participants with instructions to send them back after they have a chance to complete them. This reduces the value of the reaction sheets for two reasons. First, some, perhaps even most, of the participants will not do it. Second, the forms that are returned may not be a good indication of the reaction of the group as a whole. Therefore, have participants turn in their reaction sheets before they leave the room. If you feel that reactions would be more meaningful if participants took more time to complete them, you can send out a follow-up reaction sheet after the training together with a cover memo that says something like this: “Thanks for the reaction sheet you completed at the end of the training meeting. As you think back on the program, you may have different or additional reactions and comments. Please complete the enclosed form, and return it within the next three days. We want to provide the most practical training possible. Your feedback will help us.”

Get Honest Responses

Getting honest responses may seem to be an unnecessary requirement, but it is important. Some trainers like to know who said what. And they use an approach that lets them do just that. For example, they have the participants sign the forms. Or they tell them to complete the form and leave it at their place. In one program, the trainers used a two-sided form. One side was the reaction sheet. The other side sought attendance information: Participants were asked to give their name, department, and so on. I don’t know whether the trainers were being clever or stupid.
In some programs, like those at the University of Wisconsin Management Institute, there is space at the bottom of the reaction sheets labeled signature (optional). It is often meaningful to know who made a comment for two reasons: if the comment is positive, so you quote that person in future program brochures, or so that you can contact that person relative to the comment or suggestion.

Where people attend outside programs, they are usually free to give their honest opinion even if it is critical. They see little or no possibility of negative repercussions. The situation can be different in an in-house program. Some participants may be reluctant to make a critical reaction or comment because they fear repercussions. They may be afraid that the instructor or training department staff will feel that the reaction is not justified and there is something wrong with the participant, even that trainers might tell the participant's boss about the negative reaction and that it could affect their future. Therefore, to be sure that reactions are honest, you should not ask participants to sign the forms. Also, you should ask that completed forms be put in a pile on a table so there is no way to identify the person who completed an individual form. In cases where it would be beneficial to identify the individual, the bottom of the form can have a space for a signature that is clearly labeled as optional.

Develop Acceptable Standards

A numerical tabulation can be made of all the forms discussed and shown in this chapter. Exhibit 4.7 shows a tabulation of the reactions of twenty supervisors to the form shown in Exhibit 4.1. The following five-point scale can be used to rate the responses on a form.

Excellent = 5  Very good = 4  Good = 3  Fair = 2  Poor = 1

You tally the responses in each category for all items. For each item, you multiply the number of responses by the corresponding weighting and add the products together. Then you divide by the total number of responses received. For example, you calculate the rating for item 1, subject, as follows:

\[(10 \times 5 = 50) + (5 \times 4 = 20) + (3 \times 3 = 9) + (1 \times 2 = 2) + (1 \times 1 = 1) = 82\]

The rating is 82/20 or 4.1
Evaluating Reaction

Exhibit 4.7. Tabulating Responses to Reaction Sheets

Please give us your frank reactions and comments. They will help us to evaluate this program and improve future programs.

Leader: Tom Jones  Subject: Leadership

1. How do you rate the subject? (interest, benefits, etc.)
   - 10 Excellent  Comments and suggestions:
   - 9 Very good
   - 8 Good
   - 7 Fair
   - 6 Poor
   - Rating = 4.1

2. How do you rate the conference leader? (knowledge of subject matter, ability to communicate, etc.)
   - 8 Excellent  Comments and suggestions:
   - 7 Very good
   - 6 Good
   - 5 Fair
   - 4 Poor
   - Rating = 3.8

3. How do you rate the facilities? (comfort, convenience, etc.)
   - 7 Excellent  Comments and suggestions:
   - 6 Very good
   - 5 Good
   - 4 Fair
   - 3 Poor
   - Rating = 4.0

4. What would have improved the program?

Note: Ratings are on a five-point scale.

You can use these ratings to establish a standard of acceptable performance. This standard can be based on a realistic analysis of what can be expected considering such conditions as budgets, facilities available,
skilled instructors available, and so on. For example, at the University of Wisconsin Management Institute, the standard of subjects and leaders was placed at 4.7 on a five-point scale. This standard was based on past ratings. In this situation, budgets were favorable, and most of the instructors were full-time, professional trainers operating in nice facilities. In many organizations, limitations would lower the standard. You can have different standards for different aspects of the program. For example, the standard for instructors could be higher than the standard for facilities. The standards should be based on past experience, considering the ratings that effective instructors have received.

Measure Reactions Against Standards and Take Appropriate Action

Once realistic standards have been established, you should evaluate the various aspects of the program and compare your findings with the standards. Your evaluation should include impressions of the coordinator as well as an analysis of the reaction sheets of participants. Several approaches are possible if the standard is not met.

1. Make a change—in leaders, facilities, subject, or something else.
2. Modify the situation. If the instructor does not meet the standard, help by providing advice, new audiovisual aids, or something else.
3. Live with an unsatisfactory situation.
4. Change the standard if conditions change.

In regard to the evaluation of instructors, I once faced a situation that I’ll never forget. At the Management Institute, I selected and hired an instructor from General Electric to conduct a seminar for top management. He had a lot of experience, both of the subject and in conducting seminars both inside and outside the company. His rating was 3.3, far below our standard of 4.7. He said that we used reaction sheets and asked me to send him a summary. He also said, “Don, I know that you conduct and coordinate a lot of seminars. I would appreciate your personal comments and any suggestions for improvement.” I agreed to do it.

I enclosed a thank-you letter with a summary of the comment sheets. My thank-you tactfully offered the following suggestions, which, I in-
dictated, were based on the reaction sheets and on my own observations: “Use more examples to illustrate your points. Give the group more opportunity to ask questions. Ask your audio-visual department to prepare some professional slides and/or transparencies that will help to maintain interest and communicate.”

I waited for a thank-you for my constructive suggestions. I am still waiting, and this happened in 1969. I did hear through a mutual friend that the instructor was very unhappy with my letter. He complained that he had taken time from a busy schedule to speak at the University of Wisconsin, he didn’t take any fee or expenses, and the only thanks he had gotten was my letter. That was the last time he’d agree to be on our programs.

This example suggests that program coordinators should be very tactful in “helping” instructors by offering suggestions, especially if the instructors are members of top management within their own organization. One practical approach is to let instructors know ahead of time that reaction sheets will be used and that ratings will be compared with a standard. Instructors are usually eager to meet or beat the standard. If they don’t, most will either ask for helpful suggestions or decide that someone else should probably do the teaching in the future. This is usually good news for the training staff, who may want to make a change anyway.

Obviously, all reactions that can be tabulated should be tabulated and the ratings calculated. In regard to comments, trainers can either record all comments on a summary sheet or summarize the comments that are pertinent. Tabulations can even be made of similar comments.

**Communicate Reactions as Appropriate**

Trainers are always faced with decisions regarding the communication of reactions to programs. Obviously, if instructors want to see their reaction sheets, they should be shown them or at least a summary of the responses. Other members of the training department should certainly have access to them. The person to whom the training department reports, usually the manager of Human Resources, should be able to see them. Communicating the reactions to others depends on two factors: who wants to see them and with whom training staff want to communicate.
Regarding who wants to see them, training staff must decide whether it is appropriate. Is it only out of curiosity, or does the requester have legitimate reasons?

Regarding the desire of training staff to communicate the reactions, the question is how often the information should be communicated and in what detail. Those who make decisions about staffing, budgets, salary increases, promotions, layoffs, and so on should be informed. Also, as I suggested in Chapter 1, if there is an advisory committee, its members should be informed. If the concepts and principles described in Chapter 1 have been implemented, the reactions will be favorable, and top management will respect the training department and realize how much the organization needs it in good and bad times.

Summary

Measuring reaction is important and easy to do. It is important because the decisions of top management may be based on what they have heard about the training program. It is important to have tangible data that reactions are favorable. It is important also because the interest, attention, and motivation of participants has much to do with the learning that occurs. Still another reason why it is important is that trainees are customers, and customer satisfaction has a lot to do with repeat business.

This chapter has provided guidelines, forms, procedures, and techniques for measuring reaction effectively. Reaction is the first level in the evaluation process. It should be evaluated for all training programs. The responses to reaction sheets should be tabulated, and the results should be analyzed. The comments received from participants should be considered carefully, and programs should be modified accordingly. This measure of customer satisfaction can make or break a training department. It is only the first step, but it is an important one.

P.S. If you refer to reaction sheets as "smile" sheets, smile when you do so and hope that participants are smiling when they leave the program!
Chapter 5
Evaluating Learning

There are three things that instructors in a training program can teach: knowledge, skills, and attitudes. Measuring learning, therefore, means determining one or more of the following:

- What knowledge was learned?
- What skills were developed or improved?
- What attitudes were changed?

It is important to measure learning because no change in behavior can be expected unless one or more of these learning objectives have been accomplished. Moreover, if we were to measure behavior change (level 3) and not learning and if we found no change in behavior, the likely conclusion is that no learning took place. This conclusion may be very erroneous. The reason why no change in behavior was observed may be that the climate was preventing or discouraging, as described in Chapter 3. In these situations, learning may have taken place, and the learner may even have been anxious to change his or her behavior. But because his or her boss either prevented or discouraged the

Note: In the guidelines for levels 2, 3, and 4 no information has been given on how to use statistics. This subject is too complex to be included here. I encourage readers to consider statistical analysis. Consult people within your organization who are knowledgeable, and ask them to help you apply statistics to level 2 as well as to levels 3 and 4. Chapters 13, 14, 16, 18, 19, and 21 use statistics to determine the effectiveness of training.
trainee from applying his or her learning on the job, no change in behavior took place.

The measurement of learning is more difficult and time-consuming than the measurement of reaction. These guidelines will be helpful:

**Guidelines for Evaluating Learning**

1. Use a control group if practical.
2. Evaluate knowledge, skills, and/or attitudes both before and after the program.
3. Use a paper-and-pencil test to measure knowledge and attitudes.
4. Use a performance test to measure skills.
5. Get a 100 percent response.
6. Use the results of the evaluation to take appropriate action.

The remainder of this chapter suggests ways of implementing these guidelines.

**Use a Control Group If Practical**

The term control group will be used in levels 3 and 4 as well as here in level 2. It refers to a group that does not receive the training. The group that receives the training is called the experimental group. The purpose of using a control group is to provide better evidence that change has taken place. Any difference between the control group and the experimental group can be explained by the learning that took place because of the training program.

The phrase *whenever practical* is important for several reasons. For example, in smaller organizations, there will be a single training program in which all the supervisors are trained. In larger organizations, there are enough supervisors that you can have a control group as well as an experimental group. In this case, you must take care to be sure that the groups are equal in all significant characteristics. Otherwise, comparisons are not valid. It could be done by giving the training program only to the experimental group and comparing scores before training with scores after training for both the experimental and control groups. The control group would receive the training at a later time. The example of test scores later in this chapter will illustrate this.
Evaluate Knowledge, Skills, and/or Attitudes

The second guideline is to measure attitudes, knowledge, and/or attitudes before and after the program. The difference indicates what learning has taken place.

Evaluating Increase in Knowledge and Changes in Attitudes

If increased knowledge and/or changed attitudes is being measured, a paper-and-pencil test can be used. (This term must have been coined before ballpoint pens were invented.) I’ll use the Management Inventory on Managing Change (MIMC) described in Chapter 1 to illustrate.

Example 1 in Table 5.1 shows that the average score of the experimental group on the pretest (that is, on the test given before the program started) was 45.5 on a possible score of 65. The average score of the experimental group on the posttest (the same test given at the conclusion of the program) was 55.4—a net gain of 9.9.

Example 1 also shows that the average score of the control group on the pretest was 46.7 and that the score of the control group on the posttest was 48.2. This means that factors other than the training

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>45.5</td>
</tr>
<tr>
<td>Posttest</td>
<td>55.4</td>
</tr>
<tr>
<td>Gain</td>
<td>+9.9</td>
</tr>
<tr>
<td>Net Gain 9.9 - 1.5 = 8.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 2</td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>45.5</td>
</tr>
<tr>
<td>Posttest</td>
<td>55.4</td>
</tr>
<tr>
<td>Gain</td>
<td>+9.9</td>
</tr>
<tr>
<td>Net Gain 9.9 - 7.7 = 2.2</td>
<td></td>
</tr>
</tbody>
</table>
program caused the change. Therefore, the gain of 1.5 must be deducted from the 9.9 gain of the experimental group to show the gain resulting from the training program. The result is 8.4.

Example 2 in Table 5.1 shows a different story. The net gain for the control group between the pretest score of 46.7 and the posttest score of 54.4 is 7.7. When this difference is deducted from the 9.9 registered for the experimental group, the gain that can be attributed to the training program is only 2.2.

This comparison of total scores on the pretest and posttest is one method of measuring increased knowledge and/or changes in attitude. Another important measure involves the comparison of pretest and posttest answers to each item on the inventory or test. For example, this is item 4 of the MIMC described in Chapter 1: “If a change is going to be unpopular with your subordinates, you should proceed slowly in order to obtain acceptance.”

Table 5.2 shows that seven of the twenty-five supervisors in the experimental group agreed with item 4 on the pretest, and eighteen disagreed. It also shows that twenty agreed with it on the posttest, and five disagreed. The correct answer is Agree, so the positive gain was 11. Table 5.2 also shows the pretest and posttest responses from the control group. For it, the gain was 1. Therefore, the net gain due to the training program was 10.

Example 8 in Table 5.2 shows a different story. Item 8 states: “If you are promoted to a management job, you should make the job different than it was under your predecessor.”

Five of those in the experimental group agreed on the pretest, and twenty disagreed. On the posttest, six agreed, and nineteen disagreed. The correct answer is Agree. The net gain was 1. The figures for the control group were the same. So there was no change in attitude and/or knowledge on this item.

This evaluation of learning is important for two reasons. First, it measures the effectiveness of the instructor in increasing knowledge and/or changing attitudes. It shows how effective he or she is. If little or no learning has taken place, little or no change in behavior can be expected.

Just as important is the specific information that evaluation of learning provides. By analyzing the change in answers to individual items, the instructor can see where he or she has succeeded and where he or she has failed. If the program is going to be repeated, the in-
Table 5.2: Responses to Two Items on the Management Inventory on Managing Change

Item 4. "If a change is going to be unpopular with your subordinates, you should proceed slowly in order to obtain acceptance." (The correct answer is Agree.)

<table>
<thead>
<tr>
<th></th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
<td>Disagree</td>
</tr>
<tr>
<td>Pretest</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Posttest</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Gain</td>
<td>+13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Net Gain 13 - 1 = 12</td>
<td></td>
</tr>
</tbody>
</table>

Item 8. "If you are promoted to a management job, you should make the job different than it was under your predecessor." (The correct answer is Agree.)

<table>
<thead>
<tr>
<th></th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
<td>Disagree</td>
</tr>
<tr>
<td>Pretest</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Posttest</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Gain</td>
<td>+1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Net Gain 1 - 1 = 0</td>
<td></td>
</tr>
</tbody>
</table>

Instructor can plan other techniques and/or aids to increase the chances that learning will take place. Moreover, if follow-up sessions can be held with the same group, the things that have not been learned can become the objectives of these sessions.

These examples have illustrated how a control group can be used. In most organizations, it is not practical to have a control group, and the evaluation will include only figures for those who attended the training program.

It almost goes without saying that a standardized test can be used only to the extent that it covers the subject matter taught in the training program. When I teach, I use the various inventories that I have developed as teaching tools. Each inventory includes much of the content of the corresponding program. The same principles and techniques can and should be used with a test developed specifically for the organization. For example, MGIC, a mortgage insurer in Milwaukee, has developed an extensive test covering information that its supervisors need to know.
Much of this information is related to the specific policies, procedures, and facts of the business and organization. Some of the items are true or false, while others are multiple choice, as Exhibit 5.1 shows.

The training people have determined what the supervisors need to know. Then they have written a test covering that information. They

---

Exhibit 5.1. Sample Items from a MGIC Test to Evaluate Supervisor Knowledge

1. **T or F** When preparing a Truth-in-Lending disclosure with a financed single premium, mortgage insurance should always be disclosed for the life of the loan.

2. **T or F** GE and MGIC have the same refund policy for refundable single premiums.

3. **T or F** MGIC, GE, and PMI are the only mortgage insurers offering a non-refundable single premium.

4. _____ Which of the following is not a category in the loan program report?
   a. Loans approved
   b. Loans-in-suspense
   c. Loans denied
   d. Loans received

5. _____ Which of the following do not affect the MGIC Plus buying decision?
   a. Consumer
   b. Realtor
   c. MGIC underwriter
   d. Secondary market manager
   e. Servicing manager
   f. All the above
   g. None of the above
   h. Both b and c
   i. Both c and e

6. _____ The new risk-based capital regulations for savings and loans have caused many of them to
   a. Convert whole loans into securities
   b. Begin originating home equity loans
   c. Put MI on their uninsured 90s
   d. All the above
   e. Both e and f
   f. Both b and f
have combined true-or-false statements with multiple-choice items to make the test interesting. A tabulation of the pretest responses to each item will tell the instructors what the supervisors do and do not know before they participate in the program. It will help them to determine the need for training. If everyone knows the answer to an item before the program takes place, there is no need to cover the item in the program. A tabulation of posttest responses will tell the instructor where he or she has succeeded and where he or she has failed in getting the participants to learn the information that the test covers. It will help instructors to know what they need to emphasize and whether they need to use more aids in future programs. It will also tell them what follow-up programs are needed.

This type of test is different from the inventories described earlier. Participants must know the answers to the questions in Exhibit 5.1. Therefore, those who take the posttest put their name on it, and they are graded. Those who do not pass must take further training until they pass the test.

In regard to the inventories, there is no need to identify the responses and scores of individual persons. The scoring sheet shown in Exhibit 5.2 is given to supervisors. They score their own inventory and circle the number of each item that they answered incorrectly. They keep their inventory and turn in the scoring sheet. These can be tabulated to determine both the total score and the responses to individual items. You can then use the resulting numbers as shown in Tables 5.1 and 5.2.

Both the MIMC and the MGIC examples are typical of efforts to measure increase in knowledge and/or changes in attitudes.

### Exhibit 5.2. Scoring Sheet for the Management Inventory on Managing Change

<table>
<thead>
<tr>
<th>Management Inventory on Managing Change</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please circle by number those items you answered incorrectly according to the scoring key. Then determine your score by subtracting the number wrong from 65.</td>
<td></td>
</tr>
<tr>
<td>1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18</td>
<td></td>
</tr>
<tr>
<td>19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34</td>
<td></td>
</tr>
<tr>
<td>35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50</td>
<td></td>
</tr>
<tr>
<td>51  52  53  54  55  56  57  58  59  60  61  62  63  64  65</td>
<td></td>
</tr>
</tbody>
</table>

Score 65
Evaluating Training Programs

Evaluating Increase in Skills

If the objective of a program is to increase the skills of participants, then a performance test is needed. For example, some programs aim at improving oral communication skills. A trained instructor can evaluate the level of proficiency. Other participants may also be qualified if they have been given standards of performance. For the pretest, you can have each person give a short talk before any training has been given. The instructor can measure these talks and assign them a grade. During the program, the instructor provides principles and techniques for making an effective talk. The increase in skills can be measured for each succeeding talk that participants give. The same approach can be used to measure such skills as speaking, writing, conducting meetings, and conducting performance appraisal interviews. Chapter 13 describes such a program.

The same principles and techniques apply when technical skills, such as using a computer, making out forms, and selling, are taught. Of course, the before-and-after approach is not necessary where the learner has no previous skill. An evaluation of the skill after instruction measures the learning that has taken place.

Get a 100 Percent Response

Anything less than a 100 percent response requires a carefully designed approach to select a sample group and analyze the results statistically. It is not difficult to get everyone in the group to participate, and tabulations become simple. Tables 5.1 and 5.2 show how this can be done. It is desirable to analyze the tabulations shown in Tables 5.1 and 5.2 statistically, but in most organizations it is not necessary.

Take Appropriate Action

There is an old saying that, if the learner hasn’t learned, the teacher hasn’t taught. This is a good philosophy for each instructor to have. It is only too easy to blame a learner for not learning. How many times have we trainers said (or perhaps only thought) to someone whom we are teaching, “How many times do I have to tell you before you catch
on?" And usually the tone makes it clear that we are criticizing the
learner, not simply asking a question. Another old saying applies pretty
well to the same situation. When you point a finger at another person,
you are pointing three fingers at yourself! This saying, too, can be ap-
plied in many teaching situations.

The important point is that we are measuring our own effectiveness
as instructors when we evaluate participants' learning. If we haven't
succeeded, let's look at ourselves and ask where we have failed, not
what is the matter with the learners. And if we discover that we have
not been successful instructors, let's figure how we can be more effect-
te in the future. Sometimes the answer is simply better preparation.
Sometimes it's the use of aids that help us to maintain interest and
communicate more effectively. And sometimes the answer is to re-
place the instructor.

Summary

Evaluating learning is important. Without learning, no change in be-
behavior will occur. Sometimes, the learning objective is to increase
knowledge. Increased knowledge is relatively easy to measure by
means of a test related to the content of the program that we adminis-
ter before and after the training. If the knowledge is new, there is no
need for a pretest. But if we are teaching concepts, principles, and
techniques that trainees may already know, a pretest that we can com-
pare with a posttest is necessary.

We can measure attitudes with a paper-and-pencil test. For exam-
ple, programs on diversity in the workforce aim primarily at changing
attitudes. We can design an attitude survey that covers the attitudes we
want participants to have after taking part in the program. A compar-
ison of the results from before and after training can indicate what
changes have taken place. In such cases, it is important not to identify
learners so we can be sure that they will give honest answers, not the
answers that we want them to give.

The third thing that can be learned is skills. In these situations, a
performance test is necessary. A pretest will be necessary if it is possi-
bile that they already possess some of the skills taught. If you are teach-
ing something entirely new, then the posttest alone will measure the
extent to which they have learned the skill.
Chapter 6
Evaluating Behavior

What happens when trainees leave the classroom and return to their jobs? How much transfer of knowledge, skills, and attitudes occurs? That is what level 3 attempts to evaluate. In other words, what change in job behavior occurred because people attended a training program?

It is obvious that this question is more complicated and difficult to answer than evaluating at the first two levels. First, trainees cannot change their behavior until they have an opportunity to do so. For example, if you, the reader of this book, decide to use some of the principles and techniques that I have described, you must wait until you have a training program to evaluate. Likewise, if the training program is designed to teach a person how to conduct an effective performance appraisal interview, the trainee cannot apply the learning until an interview is held.

Second, it is impossible to predict when a change in behavior will occur. Even if a trainee has an opportunity to apply the learning, be or she may not do it immediately. In fact, change in behavior may occur at any time after the first opportunity, or it may never occur.

Third, the trainee may apply the learning to the job and come to one of the following conclusions: “I like what happened, and I plan to continue to use the new behavior.” “I don’t like what happened, and I will go back to my old behavior.” “I like what happened, but the boss and/or time restraints prevent me from continuing it.” We all hope that the rewards for changing behavior will cause the trainee to come
to the first of these conclusions. It is important, therefore, to provide help, encouragement, and rewards when the trainee returns to the job from the training class. One type of reward is intrinsic. This term refers to the inward feelings of satisfaction, pride, achievement, and happiness that can occur when the new behavior is used. Extrinsic rewards are also important. These are the rewards that come from the outside. They include praise, increased freedom and empowerment, merit pay increases, and other forms of recognition that come as the result of the change in behavior.

In regard to reaction and learning, the evaluation can and should take place immediately. When you evaluate change in behavior, you have to make some important decisions: when to evaluate, how often to evaluate, and how to evaluate. This makes it more time-consuming and difficult to do than levels 1 and 2. Here are some guidelines to follow when evaluating at level 3.

Guidelines for Evaluating Behavior

1. Use a control group if practical.
2. Allow time for behavior change to take place.
3. Evaluate both before and after the program if practical.
4. Survey and/or interview one or more of the following: trainees, their immediate supervisor, their subordinates, and others who often observe their behavior.
5. Get 100 percent response or a sampling.
6. Repeat the evaluation at appropriate times.
7. Consider cost versus benefits.

The remainder of this chapter suggests ways of implementing these guidelines.

Use a Control Group if Practical

Chapter 5 described the use of control groups in detail. A comparison of the change in behavior of a control group with the change experienced by the experimental group can add evidence that the change in behavior occurred because of the training program and not for other reasons. However, caution must be taken to be sure the two groups are
equal in all factors that could have an effect on behavior. This may be
difficult if not impossible to do.

Allow Time for Behavior Change to Take Place

As already indicated, no evaluation should be attempted until trainees
have had an opportunity to use the new behavior. Sometimes, there is
an immediate opportunity for applying it on the job. For example, if
the training program is trying to change attitudes toward certain sub-
ordinates by teaching about diversity in the workforce, participants
have an immediate opportunity to change attitudes and behavior as
soon as they return to the job. Or if the program teaches management
by walking around (MBWA), as encouraged by United Airlines and
Hewlett-Packard, participants have an opportunity to use the tech-
nique right away. However, if the purpose of the training is to teach a
foreman how to handle a grievance, no change in behavior is possible
until a grievance has been filed.

Even if a participant has an immediate opportunity to transfer the
training to the job, you should still allow some time for this transfer to
occur. For some programs, two or three months after training is a
good rule of thumb. For others, six months is more realistic. Be sure
to give trainees time to get back to the job, consider the new suggested
behavior, and try it out.

Evaluate Both Before and
After the Program if Practical

Sometimes evaluation before and after a program is practical, and
sometimes it is not even possible. For example, supervisors who attend
the University of Wisconsin Management Institute training programs
sometimes do not enroll until a day or two before the program starts.
It would not be possible for the instructors or designated research stu-
dents to measure their behavior before the program. In an in-house
program, it would be possible, but it might not be practical because of
time and budget constraints.

It is important when planning a supervisory training program to
determine the kind of behavior that supervisors should have in order
to be most effective. Before the training program, you measure the behavior of the supervisors. After the program, at a time to be determined as just outlined, you measure the behavior of the supervisors again to see whether any change has taken place in relation to the knowledge, skills, and/or attitudes that the training program taught. By comparing the behaviors observed before and after the program, you can determine any change that has taken place.

An alternative approach can also be effective. Under this approach, you measure behavior after the program only. Those whom you interview or survey are asked to identify any behavior that was different than it had been before the program. This was the approach that we used at the Management Institute to evaluate the three-day supervisory training program called Developing Supervisory Skills. Chapter 15 describes this evaluation.

In some cases, the training professionals and/or persons whom they select can observe the behavior personally.

Survey and/or Interview Persons
Who Know the Behavior

As the guideline suggests, evaluators should survey and/or interview one or more of the following: trainees, their immediate supervisor, their subordinates, and others who are knowledgeable about their behavior.

Four questions need to be answered: Who is best qualified? Who is most reliable? Who is most available? Are there any reasons why one or more of the possible candidates should not be used?

If we try to determine who is best qualified, the answer is probably the subordinates who see the behavior of the trainee on a regular basis. In some cases, others who are neither boss nor subordinate have regular contact with the trainee. And, of course, the trainee knows (or should know) his or her own behavior. Therefore, of the four candidates just named, the immediate supervisor may be the person least qualified to evaluate the trainee unless he or she spends a great deal of time with the trainee.

Who is the most reliable? The trainee may not admit that behavior has not changed. Subordinates can be biased in favor of or against the trainee and therefore give a distorted picture. In fact, anyone can give
a distorted picture, depending on his or her attitude toward the trainee or the program. This is why more than one source should be used.

Who is the most available? The answer depends on the particular situation. If interviews are to be conducted, then availability is critical. If a survey questionnaire is used, it is not important. In this case, the answer depends on who is willing to spend the time needed to complete the survey.

Are there any reasons why one or more of the possible candidates should not be used? The answer is yes. For example, asking subordinates for information on the behavior of their supervisor may not sit well with the supervisor. However, if the trainee willing to have subordinates questioned, this may be the best approach of all.

A significant decision is whether to use a questionnaire or an interview. Both have their advantages and disadvantages. The interview gives you an opportunity to get more information. The best approach is to use a patterned interview in which all interviewees are asked the same questions. Then you can tabulate the responses and gather quantitative data on behavior change.

But interviews are very time-consuming, and only a few can be conducted if the availability of the person doing the interviewing is limited. Therefore, a small sample of those trained can be interviewed. However, the sample may not be representative of the behavior change that took place in trainees. And you cannot draw conclusions about the overall change in behavior. Exhibit 6.1 shows a patterned interview that can be used as is or adapted to your particular situation.

A survey questionnaire is usually more practical. If it is designed properly, it can provide the data that you need to evaluate change in behavior. The usual problem of getting people to take the time to complete it is always present. However, you can overcome this problem by motivating the people whom you ask to complete the survey. Perhaps there can be some reward, either intrinsic or extrinsic, for doing it. Or a person can be motivated to do it as a favor to the person doing the research. Producing information for top management as the reason for doing it may convince some. If the instructor, the person doing the evaluation, or both have built a rapport with those who are asked to complete the survey, they usually will cooperate. Exhibit 6.2 shows a survey questionnaire that you can use as is or adapt to your organization.
Exhibit 6.1. Patterned Interview

The interviewer reviews the program with the interviewee and highlights the behavior that the program encouraged. The interviewer then clarifies the purpose of the interview, which is to evaluate the effectiveness of the course so that improvements can be made in the future. Specifically, the interview will determine the extent to which the suggested behaviors have been applied on the job. If they have not been applied, the interview will seek to learn why not. The interviewer makes it clear that all information will be held confidential so that the answers given can be frank and honest.

1. What specific behavior were you taught and encouraged to use?

2. When you left the program, how eager were you to change your behavior on the job?
   ____ Very eager  ____ Quite eager  ____ Not eager
   Comments:

3. How well equipped were you to do what was suggested?
   ____ Very  ____ Quite  ____ Little  ____ None

4. If you are not doing some of the things that you were encouraged and taught to do, why not?

<table>
<thead>
<tr>
<th>How Significant?</th>
<th>Very</th>
<th>To some extent</th>
<th>Not</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. It wasn't practical for my situation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. My boss discourages me from changing.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. I haven't found the time.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. I tried it, and it didn't work.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Other reasons.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. To what extent do you plan to do things differently in the future?
   ____ Large extent  ____ Some extent  ____ No extent

6. What suggestions do you have for making the program more helpful?


**Evaluating Training Programs**

Exhibit 6.2. Survey Questionnaire

**Instructions:** The purpose of this questionnaire is to determine the extent to which those who attended the recent program on leadership methods have applied the principles and techniques that they learned to the job. The results of the survey will help us to assess the effectiveness of the program and identify ways in which it can be made more practical for those who attend. Please be frank and honest in your answers. Your name is strictly optional. The only reason we ask is that we might want to follow up on your answers to get more comments and suggestions from you.

Please circle the appropriate response after each question.

5 = Much more  4 = Some more  3 = The same  2 = Some less  1 = Much less

<table>
<thead>
<tr>
<th>Time and energy spent after the program compared to time and energy spent before the program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding and Motivating</td>
</tr>
<tr>
<td>1. Getting to know my employees</td>
</tr>
<tr>
<td>2. Listening to my subordinates</td>
</tr>
<tr>
<td>3. Praise good work</td>
</tr>
<tr>
<td>4. Talking with employees about their families and other personal interests</td>
</tr>
<tr>
<td>5. Asking subordinates for their ideas</td>
</tr>
<tr>
<td>6. Managing by walking around</td>
</tr>
<tr>
<td>Orienting and Training</td>
</tr>
<tr>
<td>7. Asking new employees about their families, past experience, etc.</td>
</tr>
<tr>
<td>8. Taking new employees on a tour of the department and other facilities</td>
</tr>
<tr>
<td>9. Introducing new employees to their coworkers</td>
</tr>
<tr>
<td>10. Using the four-step method when training new and present employees</td>
</tr>
<tr>
<td>11. Being patient when employees don’t learn as fast as I think they should</td>
</tr>
<tr>
<td>12. Tactfully correcting mistakes and making suggestions</td>
</tr>
<tr>
<td>13. Using the training inventory and timetable concept</td>
</tr>
</tbody>
</table>

What would have made the program more practical and helpful to you?

Name (optional) ________________________________
Get 100 Percent Response or a Sampling

The dictum that something beats nothing can apply when you evaluate change in behavior. The person doing the evaluation can pick out a few "typical" trainees at random and interview or survey them. Or you can interview or survey the persons most likely not to change. The conclusion might be that, if Joe and Charlie have changed their behavior, then everyone has. This conclusion may or may not be true, but the approach can be practical. Obviously, the best approach is to measure the behavior change in all trainees. In most cases, this is not practical. Each organization must determine the amount of time and money that it can spend on level 3 evaluation and proceed accordingly.

Repeat the Evaluation at Appropriate Times

Some trainees may change their behavior as soon as they return to their job. Others may wait six months or a year or never change. And those who change immediately may revert to the old behavior after trying out the new behavior for a period of time. Therefore, it is important to repeat the evaluation at an appropriate time.

I wish I could describe what an appropriate time is. Each organization has to make the decision on its own, the kind of behavior, the job climate, and other significant factors unique to the situation. I would suggest waiting two or three months before conducting the first evaluation, the exact number depending on the opportunity that trainees have to use the new behavior. Perhaps another six months should elapse before the evaluation is repeated. And, depending on circumstances and the time available, a third evaluation could be made three to six months later.

Consider Cost Versus Benefits

Just as with other investments, you should compare the cost of evaluating change in behavior with the benefits that could result from the evaluation. In many organizations, much of the cost of evaluation at level 3 is in the staff time that it takes to do. And time is money. Other costs of evaluation can include the hiring of an outside expert to guide
or even conduct the evaluation. For example, I have recently been hired by Kemper Insurance, Ford, GE, Blockbuster, and Northern States Power to present and discuss the four levels of evaluation with their training staff. At Kemper, I was asked to offer specific suggestions and return three months later to comment on the evaluations that they had done. (Chapter 12 describes one of their evaluations.) In these instances, I was called in to evaluate a specific program and to provide guidelines and specific suggestions on how programs could be evaluated at all four levels. Other consultants can be called in to evaluate the changes in behavior that result from a specific program. You should consider such costs as these when you decide whether to evaluate changes in behavior.

The other factor to consider is the benefits that can be derived from evaluation including changes in behavior and final results. The greater the potential benefits, the more time and money can be spent on the evaluation not only of behavior change but also in level 4 also. Another important consideration is the number of times the program will be offered. If it is run only once and it will not be repeated, there is little justification for spending time and money to evaluate possible changes in behavior. However, if a program is going to be repeated, the time and money spent evaluating it can be justified by the possible improvements in future programs.

It is important to understand that change in behavior is not an end in itself. Rather, it is a means to an end: the final results that can be achieved if change in behavior occurs. If no change in behavior occurs, then no improved results can occur. At the same time, even if change in behavior does occur, positive results may not be achieved. A good example is the principle and technique of managing by walking around (MBWA). Some organizations, including United Airlines and Hewlett-Packard, have found that higher morale and increased productivity can result. These organizations therefore encourage managers at all levels to walk among the lowest-level employees to show that they care. Picture a manager who has never shown concern for people. He attends a seminar at which he is told to change his behavior by walking around among lower-level employees to show that he cares. So the manager—for the first time—changes his behavior. He asks one employee about the kids. He comments to another employee regarding a vacation trip that the employee's family is planning. And he asks another employee
about Sam, the pet dog. (The manager has learned about these things before talking to the three employees.) What are the chances that the three employees are now going to be motivated to increase their productivity because the manager really cares? Or will they look with suspicion on the new behavior and wonder what the boss is up to? The manager’s change in behavior could even have negative results. This possibility underlines the fact that some behavior encouraged in the classroom is not appropriate for all participants. Encouraging supervisors to empower employees is a behavior that would not be appropriate in departments that had a lot of new employees, employees with negative attitudes, or employees with limited knowledge.

**Summary**

Level 3 evaluation determines the extent to which change in behavior occurs because of the training program. No final results can be expected unless a positive change in behavior occurs. Therefore, it is important to see whether the knowledge, skills, and/or attitudes learned in the program transfer to the job. The process of evaluating is complicated and often difficult to do. You have to decide whether to use interviews, survey questionnaires, or both. You must also decide whom to contact for the evaluation.

Two other difficult decisions are when and how often to conduct the evaluation. Whether to use a control group is still another important consideration. The sum of these factors discourages most trainers from even making an attempt to evaluate at level 3. But something beats nothing, and I encourage trainers to do some evaluating of behavior even if it isn’t elaborate or scientific. Simply ask a few people, Are you doing anything different on the job because you attended the training program?

If the answer is yes, ask, Can you briefly describe what you are doing and how it is working out? If you are not doing anything different, can you tell me why? Is it because you didn’t learn anything that you can use on the job? Does your boss encourage you to try out new things, or does your boss discourage any change in your behavior? Do you plan to change some of your behavior in the future? If the answer is yes, ask, What do you plan to do differently?
Questions like these can be asked on a questionnaire or in an interview. A tabulation of the responses can provide a good indication of changes in behavior.

If the program is going to be offered a number of times in the future and the potential results of behavior changes are significant, then a more systematic and extensive approach should be used. The guidelines in this chapter will prove helpful.
Appendix 2
Designing questionnaires and analysing the data

Most of the information used by evaluators is gathered by the use of structured interviews and questionnaires. There is a good deal of similarity between the two methods and the two techniques can often be combined with the evaluator administering a questionnaire on the more quantitative aspects, then following this up with an interview.

Questionnaires are used more frequently than interviews; this is due to a combination of some of the following advantages:

- Questionnaires are much cheaper; it is expensive to have interviewers travelling long distances and interviewing large numbers of people.
- Much larger samples can be taken using questionnaires and the questions can be administered to a large sample simultaneously.
- By careful design, the processing of questionnaire answers can be made very simple and efficient.
- It is often easier to convince respondents of the anonymity of their answers if they are filling in a questionnaire as opposed to undergoing a face to face interview.

On the other hand, questionnaires have a number of disadvantages when compared with interviews. These stem primarily from the greater flexibility within the interview situation, where the evaluator can follow leads as necessary and is not confined to the printed questions. Questionnaires are also likely to elicit response biases as respondents have a tendency to answer the questions in what they perceive to be a socially acceptable manner.

Planning the questionnaire

The most important stage in the use of a questionnaire is the planning before it is drafted. The following questions need to be answered:

- Is a questionnaire the best method of collecting the data?
- What information is required?
- Who is to provide this information?
- What type of analysis will be carried out on the information collected?

All of these questions are, of course, interlinked and decisions taken on one question may well determine the answer to other questions.
Appendices

The type of information required will often determine the format of the questionnaire and this will control the analysis of the data collected. It is therefore necessary to think about this before the design stage.

The ideal sample consists of everyone who has relevant information and one of the advantages of a questionnaire approach is that this is sometimes possible. If the numbers involved are less than 200, this is probably the best strategy. If very large numbers are involved or the resources available to collect and analyse the data are limited, it may be necessary to select a sample from the total population (i.e., who is of interest). If the sample is carefully drawn, it should be possible to use the data as representative of that which would have been collected had the whole population responded.

A simple random sample of respondents can be selected by procedures such as drawing numbers from a hat, taking names at regular intervals from an alphabetical list, or using tables of random numbers. If the variation of opinions within the total population is not thought to be great, a simple random sample of 20 per cent to 30 per cent should give representative information.

If it is thought likely that there will be wide variations in the opinion held by people in different parts of the organization or in different organizations, then a stratified sample may give a better estimate. In this, the total population is broken down into major divisions or strata, and a random sample is taken from each stratum. For instance, in following up a junior management programme, the strata could be the functions in which the people now work. Some strata will have more people in them and need a bigger sample; some will have more variation in them and again need a bigger sample.

A questionnaire can be completed either with or without supervision. The method of administration will affect the design and must be taken into account during the planning stage. Unsupervised questionnaires need very careful design and pilot runs will be necessary to eliminate ambiguities. They must also be simple. Supervised questionnaires can be more complex.

Closed questions, where the respondent is asked to select one answer from a number of alternatives, are easy to analyse and the questionnaire can be designed so that the analysis may be carried out mechanically. If the sample is large and a number of issues are to be investigated then most of the questions should be of this type. Open-ended questions, where respondents are allowed to write in whatever they please, are very difficult to analyse. It will be necessary to establish categories so that the information can be summarized in a usable form, and this will involve a good deal of time. This must be clearly understood and anticipated at the planning stage. Open questions allow people the opportunity to express their particular point of view rather than being confined to pre-determined answers. This makes some people feel more at ease. The choice of format for the questions should be governed by the sort of information required, the ease with which useful information can be extracted from the responses, and convenience for the respondents. The co-
Questionnaire construction

To ensure willing co-operation, the purpose of the questionnaire must be explained either in a written introduction or by the person administering it. Instructions on how to complete it should be simple and clear. Above all, the document should give the impression that it has been carefully prepared and produced.

Answers to early questions tend to be unreliable, so it is best to start with something factual like personal details. If response becomes mechanical, the answers again become inaccurate. It is therefore worth thinking about different sections to the document, each with a different layout.

All questions should be written in a way which helps those who are answering them to do so accurately. The questions should therefore be as short as possible and be phrased in language which the respondents will understand. The intention is not to confuse them with complicated constructions; even the use of negatives, as in the first part of this sentence, will confuse some.

Closed questions will often take up a large part of the questionnaire as they are easier for the respondent to answer and for the evaluator to analyse. The simplest form is a binary question like:

In your present job is it necessary for you to:

- Diagnose mechanical faults in __________ Yes/No
- Repair _________________ Yes/No
- Supervise someone using __________ Yes/No

Where there are a number of possible answers, a polylog question can be used:

To what extent are you involved in writing proposals for ________________?

- I write them
- I advise on them
- I make some recommendations
- I am not involved

Sometimes preference scales are used to assess the strength of attitudes. There are several formats for this. The best offer two opposite statements and a space for responses.

I find my work very interesting 1 2 3 4 5 6 7 I find my work deadly dull

The response space can be labelled rather than numbered.
Appendices

<table>
<thead>
<tr>
<th>Statement X</th>
<th>Strongly agree X</th>
<th>Agree X</th>
<th>Neutral</th>
<th>Agree Y</th>
<th>Strongly agree Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>The interviewer seemed very interested in me as a person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When it is difficult to produce opposites an agree/disagree format can be used:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel that...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...is an important objective of my job</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Open-ended questions can often be included at the pilot stage and later converted into questions which are easier to analyse. The range of responses which are offered can be content analysed and the themes then used as categories for closed questions. For example, in the pilot we might ask, 'What topics which would be useful to you in your job, were not covered on the course?'. The question which would then be included in the questionnaire proper might be:

The following have been recommended as areas where extensions to the basic course would be helpful:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Do you agree?</th>
<th>If yes, what particularly?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>ronework analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The question would usually be followed by an open-ended one:

Are there other topics which you think should be included?

Topic... What would you have found particularly useful?

The value in this sort of approach is that it increases the response rate. Most people will answer 'yes' or 'no' if asked, 'Do you think that X would have been useful to you?'. Not many people answer questions of the type, 'What do you think ought to be done?'.

The construction of the questionnaire is not complete until it has been tried on a sample of the target population and shown to give the sorts of answers which were expected. Respondents often do not interpret questions in the same way as the writer and the only way to sort out ambiguities is to ask the questions and discuss the answers. The best
way to carry out the pilot is to sit with a few of the respondents and encourage them to discuss questions which are difficult to understand or to answer.

Postal distribution is the most popular method but contacting the respondents personally or through an agent is likely to produce a better response rate. As well as the questionnaire itself, there will usually be a covering letter. This should explain the purpose of the investigation and thank the respondent for the time spent in answering the questions. If there is an official sponsor of the project this should also be stated. The date by which the questionnaire is to be returned should be stated. Don't give respondents more than three weeks or they will put it to one side and forget it. If the form is to be returned by post, a stamped addressed envelope should be provided.

The intention is to get as high a response rate as possible. When less than 70 per cent of the questionnaires are returned, there must be grave doubts about whether the responses are representative. Random sampling is not achieved by low response rates.

The simplest method of summarizing questionnaire data is some form of frequency statement. For example:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree X</th>
<th>Agree X</th>
<th>Undecided</th>
<th>Agree Y</th>
<th>Strongly agree Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>3</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

This represents the distribution of the opinions of 50 people on the statements X and Y.

A refinement would be to express the numbers in the form as percentages of the total number surveyed. In our example this would become:

<table>
<thead>
<tr>
<th>SA X</th>
<th>A X</th>
<th>Un</th>
<th>A Y</th>
<th>SA Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>6%</td>
<td>50%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

This is easier to understand when the number of people surveyed is not a round number. Percentages can be misleading if small numbers are involved. The total number surveyed should appear somewhere in the summary. It is, of course, easy to show frequencies or percentages as bar graphs if it is felt that this is likely to increase understanding.

Sometimes the opinions are reduced to a mean response for the purpose of comparisons. The boxes are given numbers and the frequencies multiplied by these. For instance, in the example above:

<table>
<thead>
<tr>
<th>X</th>
<th>3</th>
<th>25</th>
<th>10</th>
<th>10</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>3 + 50 + 50 + 10 + 10</td>
<td>133</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendices

This is a very dubious exercise. There is no reason to believe that the intervals between the boxes are equal and there is therefore no justification for using processes of multiplication or division.

If it is intended to compare one set of opinions with another, the correct method is to use the chi-squared statistic rather than the mean response. Suppose that we have two sets of opinions extracted from two courses about a particular issue.

<table>
<thead>
<tr>
<th>Statement X</th>
<th>Agree X</th>
<th>Tend to agree X</th>
<th>Undecided</th>
<th>Tend to agree Y</th>
<th>Agree Y</th>
<th>Statement Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course A</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Course B</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

If the proportions across the boxes are similar then there is no statistical difference between the frequencies. In this case we could estimate the frequency in any one box from the frequencies in the others and produce a figure that is quite close to the one actually found.

Estimating the frequencies in the boxes and comparing these with the actual frequencies found is the basis of the chi-squared test. Where large differences are found, the frequencies can be said to represent different opinions, or pass rates or whatever.

To take the example above. First of all find the row and column totals

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>N</th>
<th>Y</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course A</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Course B</td>
<td>5</td>
<td>10</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>Totals</td>
<td>15</td>
<td>30</td>
<td>35</td>
<td>100</td>
</tr>
</tbody>
</table>

Now calculate the 'expected value' in each cell from the formula:

Row total × Column total

Overall total

For the first cell this is

\[
\frac{50 \times 15}{100} = 7.5
\]

For the second

\[
\frac{50 \times 30}{100} = 15
\]

For the third

\[
\frac{50 \times 35}{100} = 17.5
\]
Appendix 2: Designing questionnaires and analyzing the data

Fill in the new block by writing the expected values in brackets underneath the actual, observed values.

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>10</th>
<th>5</th>
<th>5</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>(7.5)</td>
<td>(15)</td>
<td>(17.5)</td>
<td>(5)</td>
<td>(5)</td>
<td>(5)</td>
<td>(50)</td>
</tr>
<tr>
<td>(7.5)</td>
<td>(15)</td>
<td>(17.5)</td>
<td>(5)</td>
<td>(5)</td>
<td>(5)</td>
<td>(50)</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>35</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Subtract all the expected values from the observed values and calculate chi-squared from the formula

\[ \chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \]

i.e., \[ \chi^2 = \frac{(10 - 7.5)^2}{7.5} + \frac{(20 - 15)^2}{15} + \frac{(10 - 17.5)^2}{17.5} \]

\[ + \frac{(5 - 5)^2}{5} + \frac{(5 - 5)^2}{5} + \frac{(5 - 7.5)^2}{7.5} + \frac{(10 - 15)^2}{15} \]

\[ + \frac{(25 - 17.5)^2}{17.5} + \frac{(5 - 5)^2}{5} + \frac{(5 - 5)^2}{5} \]

\[ = 0.833 + 1.667 + 3.21 + 0 + 0 + 0.833 + 1.667 + 3.21 + 0 + 0 \]

\[ \chi^2 = 11.42 \]

The figure of 11.42 is a measure of the difference in the opinions expressed by the two courses. We must now decide whether it is large enough to discount chance variation and state what the opinions are.

The table of critical values (Table A2.1) is based on the 1 in 20 criterion. The degrees of freedom are calculated from:

\[ df = (\text{Number of rows} - 1) (\text{Number of columns} - 1) \]

In this case we have

\[ df = (2 - 1) (5 - 1) = 4 \]

The critical value for df 4 is 9.49. The figure for \( \chi^2 \) that we have calculated (11.42) is larger than this thus we can say that the expressed opinions of the two groups are different.
Table A2.1  Critical values for $\chi^2$ ($p < 0.05$)

The value for $\chi^2$ is computed from

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$
Appendix 3  Designing interviews and analysing the data

The interview is a widely used technique for gathering evaluation data. The interviewer can ask direct questions, and further probing and clarification is possible as the interview proceeds. This flexibility is very valuable for exploring issues as it can give more depth to the investigation than is possible when using questionnaires.

Interviews may be highly structured, resembling questionnaires, but usually start with general questions to allow the respondent to talk about some of the issues which he or she feels are important. Sometimes interviews will be exploratory and will have very little pre-planned structure. An example would be the rather informal discussions of 'how things are going' which take place in the bar on residential courses.

Interviewing typically involves a one-to-one interaction but it can be carried out with a group. Group interviews save time and allow the respondents to build on each other's responses, but this influence sometimes leads the group in a direction which none of the individuals would have chosen. The situation may also inhibit the contribution of group members as some people are inclined not to express views if they feel that they are in a minority. The empathic relationship which is the hallmark of a good interview is more difficult to achieve in a group setting. There is also the problem of the situation giving undue prominence to the statements of those who are more articulate or more confident.

Planning

The first decision to be made by the evaluator is whether interviewing is the most appropriate method of data collection. It will be particularly useful when: re-appraising previously identified training needs; exploring the extent of transfer of learning; examining the effectiveness of particular training methods; and, when trying to relate activities to organizational goals and purposes.

A major drawback of interviews is the time taken to conduct and analyze them. As we found with questionnaires, personal bias can distort the data. With interviews this is not only self-report bias, but also the
bias of the interviewer. The type of question asked and the nature of the interaction will encourage certain kinds of responses and discourage others. Interviewing takes considerable skill if valid data is to be collected. Interviewers must understand their own biases and those of the respondents. They must also be able to listen actively and to change the shape of the interview in order to probe issues which arise. If the respondents are to raise sensitive issues and offer frank statements, it will also be necessary to establish an empathic relationship with the respondents.

The next step in planning is to decide what information is sought and thus what questions must be asked. A rough interview schedule is drafted and tested on a colleague. It is then refined and piloted with a few members of the target population. The schedule will help the interviewer by providing a reminder of the points to be explored. It should not be so detailed that it dictates the whole pattern of the interview. Interviewing should be a flexible process which allows the exploration of themes which were not anticipated when the schedule was drawn up. If it has a rigid format, the data can be more economically collected by using a questionnaire.

Questions for the schedule could include some of the following:

**Questions**

- What did you hope to get out of the course before you went on it?
- Did it meet with your expectations?
- What were the most useful things that you learned?
- What are you doing differently since the programme?
- * anything else?
- Did you talk to your supervisor when you returned?

**Probing**

- To find out if expectations were realistic.
- If not, why not?
- Useful for job performance?
- In what way useful?
- Ask for specific examples; try to connect to learning.
- Describe a specific incident.
- What kind of debrief?
- What benefit from it?

Specific questions can be asked about particular aspects of the programme; ask what candidates thought were its strengths or weaknesses; ask about aspects which were new or had been tried for the first time, etc:

- Is there anything else that you would like to say about the programme?
- Is there anything that we haven’t talked about that you think we should have talked about?

**Contracting**

Establishing a good relationship with the interviewee is the purpose of the early part of the interview. What should happen is that a form of
Data gathering

The interview will often fall into two parts: an initial exploratory phase of rather general discussion, and a second phase during which specific issues are pursued. This order is recommended as it makes it more likely that the interviewee will raise issues rather than being confined only to the areas which he or she thinks that the interviewer is interested in.

Recording information with a highly structured interview is a simple process of making short notes in the spaces left on the schedule. Less structured interviews pose more difficult problems. Taking notes is the most common method and, with practice, this will record most of the useful information. Key words, phrases and quotes are recorded during the interview and these are expanded before the next interview can interfere with the memory. Taking detailed notes during the interview will interfere with the flow and with the rapport. Few interviewees enjoy talking to the top of someone’s head.

It is possible to use a tape recorder; this has the advantage of providing a complete record of the interview. However, this procedure will inhibit some interviewees. Many people are wary of making statements on sensitive issues when these can be played back verbatim somewhere else. A further disadvantage of this method is the length of time taken to access the information. For an interview lasting one hour, it will take about two hours to extract the main points from the recording and about four hours to make a complete transcription.

While carrying out the interview, the sensitive interviewer will realize the importance of his or her own behaviour in controlling that of the interviewee. For instance, people who avoid eye contact and regularly avert their gaze when speaking are usually suspected of being ‘economical with the truth’. However, such gaze aversion can easily be induced by an interviewer who sits too close to the interviewee. Simi-
Appendices

lardy, people who make false starts to sentences and then rephrase what they were going to say, are often suspected of embroidering the story. Such behaviour can easily be induced by an interviewer who has too much eye contact and who is thus perceived to be an interrogator.

Talking is often seen as active and listening as being passive. However, effective listening is an active combination of hearing, checking understanding, clarifying contradictions and summarizing what has been said. It also requires some commitment to exploring the respondent's viewpoint in as unbiased a way as possible. We all have pre-conceived ideas based upon experience, personal values, expectations of other people and untested prejudices. All of these can filter and distort what is heard. The more aware the interviewer is of these elements, the better able he or she will be to control biases in what is recorded. Reflective summaries provide an opportunity to check and correct distortions of the messages being offered. Many programmes designed to train interviewers use video-recordings for feedback. This is an area where such feedback can be particularly valuable, allowing people to see for themselves just how biased they are.

Termination
Towards the end of the interview it is good practice to briefly restate some of the main themes and give the interviewee an opportunity to add additional comments. Some open-ended questions such as, 'Is there anything that you think I should have asked about, but haven't?', will sometimes uncover topics which have been overlooked. A little caution is in order here as some interviewees will take this opportunity to open floodgates. It may be worth while prefacing the question with, 'I'd like to use the last few minutes . . . .

The interviewee should be allowed the opportunity to ask questions. There may, for instance, still be some doubts about the purpose of the interview or the level of confidentiality. The interview should be ended properly by expressing thanks. This carries through the good atmosphere in which the interview should have taken place and a statement of the respondent's contribution helps to create goodwill.

Data analysis
Interviews can yield large amounts of information and this will need summarizing. This can be done by extracting short quotations which are thought to be representative, or by writing a short passage which is a summary of the main themes which were discussed. More often, a number of interviews are summarized for a report which will have main themes. The summary then becomes a mixture of statistical statements and more qualitative information.

Forty-three of the 57 managers interviewed had not had a debriefing session with their supervisor on return from the programme. In the 14 cases where debriefing had occurred, the benefits were reported as being:

- Creation of opportunities to use new skills (9 cases)
- Negotiation of the possibilities for progressing the action plan (6 cases)
- More open relationship with the supervisor (4 cases)
- Opportunity to discuss further development (3 cases)
Appendix 3: Designing interviews and analysing the data

Direct quotations have great impact and sometimes will be the only part which the reader remembers. Therefore, they should be used sparingly and more often when they represent a widely held view. A well-turned phrase which represents the view of only one respondent may distort the understanding of the report.
Appendix 4  Observing as an evaluative technique

One of the most direct methods of collecting evaluative data is by observing people in their work setting. The observation may be unstructured, with the person who is observing being as open-minded as possible and using his or her judgement about which events are considered important. Alternatively it may be highly structured by the use of coded schedules which guide attention to specific types of event. The latter is more likely in evaluations of training. The categories which are selected will be those where changes are expected as a result of training, or those which are thought to be particularly important to the success of the job. In practice, the observer will often use both of these approaches, as an open-ended method can complement the rather narrow field observed with a highly structured one. Sometimes the observer will start with a relatively unstructured approach and later focus on aspects which seem to be of importance. This is the approach which Parlett and Hamilton (1977) have called 'progressive focusing'.

The recording of the information may be done during the observation or immediately after it. The former is better for detail and the latter for overall impressions. The longer the interval between the observation and the recording, the less accurate the information will be. If it is intended to quote actual statements these must be recorded faithfully and in quotation marks. There may be a case for using a tape recorder to ensure accuracy in such cases. Tape recorders may well introduce problems into the situation because those being observed are less likely to act naturally when they know that they are being recorded. People do get used to them and after a while, forget that they are there, but this may not happen in a short session.

Observations are free from the biases of self-reports in interviews or questionnaires because the evaluator is directly connected with behaviour rather than someone’s perceptions of it. However, the evaluator must be sensitive to the situation and the likelihood that his or her presence will distort the performance being observed. Also the same hazards apply as in interviewing if the observer is biased and sees only what he or she expects or wants to see.
Appendix 4: Observing as an evaluative technique

Observing interpersonal skills of individuals

If used correctly, observations can be particularly helpful in examining interpersonal skills and relationships with others. In observing the interactions of an individual with colleagues or customers, a set of categories like those developed by Rackham and Morgan (1977) can be used. The thirteen types of behaviour are listed on a sheet of paper and the frequency with which they are used in some significant period of time is recorded. Specific instances of appropriate or inappropriate use of a particular category are also recorded to be used as feedback. The categories are shown in Figure 7.2. After familiarization with the use of these, the observer could work with a sheet like Table A4.1.

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Frequency</th>
<th>Specific Incidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Building</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disagreeing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defending/attacking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blocking/difficulty stating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open behaviour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing understanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seeking information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giving information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shutting out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bringing in</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


It is common experience that two people observing the same event will later give different accounts of it. Putting in a lot of structure by using a checklist like the one above, will control some of this, but it will still be necessary to practise in order to produce reliable observations.

Observing groups at work

Interpersonal relationships are a key component of working in groups and observing interactions within the group setting can provide useful information about the nature of those relationships. Broad categories like the following will reveal many of the important aspects:

- **Interruptions**
  - Who interrupts the most, and the least?
  - Who is interrupted the most, and the least?

- **Air space**
  - Who talks most, and who least?
  - Who attempts to dominate the conversation?

- **Disagreement**
  - How often do they disagree?
  - How do they settle disagreements?
Support

- How often do they support each other?
- Who supports whom?

If the purpose of the group is to make decisions, a checklist like Table A4.2 might be useful.

Table A4.2  Decision-making in groups

<table>
<thead>
<tr>
<th>Who provided the structure?</th>
<th>Group member</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Follows the structure provided by others</td>
<td>A</td>
</tr>
<tr>
<td>• Provided a plan for meeting goals which was discussed</td>
<td></td>
</tr>
<tr>
<td>• Provided a plan for reaching decisions which was implemented</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How was information given?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Gave incorrect information or withheld something important</td>
<td></td>
</tr>
<tr>
<td>• Gave information in a disorganized fashion</td>
<td></td>
</tr>
<tr>
<td>• Gave information which was relevant and concise</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who made the decisions?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Avoided making decisions, accepted others' decisions</td>
<td></td>
</tr>
<tr>
<td>• Proposed solutions for others to approve (or disapprove)</td>
<td></td>
</tr>
<tr>
<td>• Proposed the solutions which were implemented</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Investment of energy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Low—kept silent except when asked something</td>
<td></td>
</tr>
<tr>
<td>• Moderate—active for much of the time</td>
<td></td>
</tr>
<tr>
<td>• High—the most active group member</td>
<td></td>
</tr>
</tbody>
</table>

Observing a training session

One important use of observation within an evaluative strategy is the observation of training sessions. This has three main phases: some discussion with the trainer before the session, the observation itself, and some feedback afterwards.

During the pre-observation discussion, the trainer is asked what he or she is trying to achieve, i.e. for aims and objectives. The trainer is then asked to describe the shape of the session and why the specific methods...
have been chosen. Some discussion should occur about how the observational record will be used, to what extent the information is confidential, and to whom. It is also good practice to ask the trainer if there are any specific aspects of the session on which he or she would like feedback.

During the session the observer should be as unobtrusive as possible. Sometimes observers become active participants in the session, but this can be difficult for the trainer and it will also make it very difficult for the observer to record information. Notes should be made during the session with some detail on specific incidents. Some framework will usually be necessary in order to classify incidents. One which we have found useful is shown in Table A4.3 (see p 138). Under 'further comments' the following questions might be addressed:

- To what extent did the methods seem suitable for this group?
- What form of assessment of trainee progress was being used?
- How was feedback given to the trainees?

The post-observation feedback should occur as soon as possible after the session. This should begin by asking the trainer what he or she thought of the actual session compared with that expected and planned. The observer should then discuss specific incidents and how they were seen by observer, trainees and trainer. This works best when incidents where things went well are discussed first. People are generally more inclined to accept the observer's opinion when discussion of positive events precedes that of incidents when things did not go well (Stone, Guertal and McIntosh, 1984). Whatever the purpose of the observation, the feedback is intended to be a helpful reflection for the trainer and it should not be heavily judgemental.

A video camera in the background can support observations. This has the advantage of full data which can later be observed by more than one person and thus be impartially classified. This is often useful for feedback during the training of interviewers, trainers, etc. as they can afterwards see how they appear to others. It may also be suitable for research and for training in observation skills. It is time consuming and may not be a cost-effective method of gathering evaluative data.
### Table A4.3  Trainer appraisal form

<table>
<thead>
<tr>
<th>Signposting</th>
<th>X</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduced subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Referred to objectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicated main stages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summarized to consolidate stages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained procedures</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Use of aids                  |   |          |
| Flip chart                   |   |          |
| White board                  |   |          |
| Over-head projector          |   |          |
| Computer demonstration       |   |          |
| Other                        |   |          |

| Delivery technique           |   |          |
| Voice (volume, tone, pace)   |   |          |
| Listened                     |   |          |
| Use of questions             |   |          |
| Group involvement            |   |          |
| Checked understanding        |   |          |
| Eye contact                  |   |          |
| Mannerisms                   |   |          |

| Control                      |   |          |
| Allocated time well          |   |          |
| Maintained good pace         |   |          |
| Kept to subject              |   |          |

| Overall impression           |   |          |
| Knowledgeable                |   |          |
| Enthusiastic                 |   |          |
| Aware of group needs         |   |          |
| Created interest             |   |          |

Any further comments:

Observer:  
Date:
Evaluating Training Systems

Earlier chapters have focused on the design and development of the training system. Now the total system must be subjected to tryout and evaluation to determine its effectiveness and the effectiveness of each of its components. Although conclusive proof of the adequacy of the system can be obtained only by follow-up and evaluation of graduates on the job, observation of the system in action can provide valuable data for improving all parts of the system. This chapter is limited to identifying precisely what is to be evaluated, who should do the evaluating, and when and how the evaluation is to be done. It is in the next chapter that we take up the ultimate means of evaluating the effectiveness of an ongoing training system: following up graduates.

Upon completing this chapter, the reader will be able to perform as follows:

- **Behavior**: Evaluate all components of an ongoing training or development system.
- **Conditions**: Given: appropriate guidelines for conducting observation, data-gathering instruments (rating scales, questionnaires, and schedules), an ongoing training or development system, access to trainees and instructors, and clerical assistance.
- **Criterion**: In accordance with the procedures defined in this chapter.

**Approaches to Evaluation**

Several approaches to evaluation are widely used in training and development. In fact, most organizations use more than one, and some use as many as four.

**Immediate Evaluation**

Also called internal evaluation, immediate evaluation focuses on ongoing programs and activities. Its objective is to assess the quality of training programs and services while they are in operation to provide supervisors and employees or instructors and trainees...
Evaluating Training Systems

the feedback they need to improve their performance. There are three subcategories of immediate evaluation: process evaluation, participant reaction, and measurement of behavioral and attitudinal change.

Process Evaluation. Assessment that focuses on ongoing training programs—use of resources, trainee-instructor interaction, training activities, learning strategies, and so on—process evaluation asks the question, “Is the training process effective and efficient?” It makes use of observation, interview, surveys, review of test results, and reports of instructors.

Participant Reaction. Participant reaction is the most common, easiest, and most casual and subjective method of evaluating training and development programs. Feelings and opinions of trainees about the training received are collected by means of questionnaires or interviews. It addresses the question, “Did the trainees like the training?” Although not entirely lacking in merit, it is neither reliable nor valid. Results do not correlate highly with job performance or the amount and kind of resulting organizational change.

Measurement of Change. Here the focus is on the amount and kind of behavioral or attitudinal change that occurs immediately as a consequence of the training provided. Here the root question is, “Did the trainees learn?” The results of training are determined by paper-and-pencil and performance tests, attitudinal surveys, observation, and structured and unstructured interviews.

Intermediate Evaluation

Intermediate evaluation assesses trainees’ reactions, learning, and performance at the conclusion of training. It may take either of three forms: individual evaluation, organizational evaluation, and the training self-audit.

Individual Evaluation. Individual evaluation involves examining changes in on-the-job behavior of managers, supervisors, workers, clients, or customers, immediately following training, development, or other form of intervention. Again, the question is, “Are the trainees using what they learned?” Techniques include questionnaires, rating scales, interviews, tests, and observation.

Organizational Evaluation. Organizational evaluation examines organizational change immediately following the intervention in such areas as quality of communications, customer relations, job satisfaction, motivation and morale, and teamwork. The evaluation centers on the question, “Was there an immediate and observable or measurable change in the organization?” Tools include observation, questionnaires, interviews, ratings, and review of records.

Training Self-Audit. This technique is used to appraise the quality of the entire organization or any one of its elements, functions, services, or programs and to institute changes to improve them. The question here is, “Is training conducted in accordance with established standards?” Using a specific set of standards, the audit examines every major aspect of training management, organization, programs, services, facilities, resources, and outcomes and assesses the relevance and adequacy of all resources invested
in training and development—buildings, facilities, equipment, materials, instructors, support personnel, and so on.

**Ultimate Evaluation**

Also called either impact evaluation or results evaluation, ultimate evaluation assesses the long-term changes in the bottom-line results achieved by organizations, departments, managers, supervisors, or employees after they have experienced some training and development intervention. Here the question is, “Did the training make a difference in terms of the attainment of organizational goals and objectives?” Measures focus on such outcomes as increased productivity and return on investment, improved quality of products or services, and reduced operating costs. It makes use of all types of performance indicators and measures. Chapter 21 addresses ultimate evaluation.

**The Objectives of Evaluation**

**Primary Objective**

The primary and overriding objective of a program of internal evaluation is to collect data that will serve as a valid basis for improving the training or development system and maintaining quality control over its components. It must be emphasized that all components of the system and their interaction are the objects of scrutiny.

**Secondary Objectives**

There are three secondary objectives that an internal evaluation system can serve:

1. To ensure that instruction is conducted in a manner consistent with the system as planned and designed. There must be some means of ensuring that the system as observed is the same as that planned; radical departures from the planned system cannot be tolerated because the validity and effectiveness of the system cannot be determined under such circumstances.

2. To provide a basis for instructor in-service training and upgrading. An effective in-service training or upgrading program for instructors cannot be based on conjecture about deficiencies; it must be based on observed needs. Data collected by qualified and trained observers can serve well as a means of identifying areas in which additional or remedial training for instructors can be planned and implemented.

3. To provide data that are usable in effecting revisions of the instructor training course. Objective evaluation often reveals common deficiencies among the instructor force that can be prevented by introducing changes in the instructor training program.

**Components to Be Evaluated**

A learning situation involves trainees, instructors, course content, sequence, time allocations, instructional strategies, materials, equipment, and facilities. If any one of these components is substandard, the training or development program cannot be optimally effective in achieving the desired results.
Trainees

Regardless of the insight and skill with which a training system has been designed, it will fail to achieve the desired results unless the trainee input matches the target population of the system. All training and development systems have trainee prerequisites, although it must be noted that prerequisites are established initially on the basis of judgment and must be validated. If trainees do not possess the prerequisite aptitudes, skills, and backgrounds of training and experience (or if mistakes were made in establishing prerequisites), they are not ready for the system as designed. They will not be able to acquire the job skills the system aims to develop—at least not with the materials selected and within the established time limits. If it is discovered that the typical trainee does not measure up to the prerequisites set for a particular training system, changes must be made in the system to accommodate a lower-quality input, or the prerequisites themselves must be changed. Close study and observation of the pilot groups are therefore necessary to determine the compatibility of the input and the training system as designed.

Instructors

Even the best instructional strategy, facilities, equipment, and materials will fail to produce the required job skills if untrained instructors, or instructors without the required skills, implement the system. Again, it must be noted that prerequisite knowledge and skills were established by judgmental procedures and must be validated. Instructors are one of the key components of a training or development system. They must possess the required technical and pedagogical knowledge and skills, and be highly motivated, if they are to be successful in using the strategies, materials, and equipment selected or produced during the system planning phase. By observing instructors as they work with the other system components, it is possible to identify mistakes in establishing instructor prerequisites. At the very least it will be possible to identify instances where the faulty component is the instructor.

Course Content

The selection of course content, earlier, is largely a judgmental procedure. Although in the final analysis the adequacy of content taught, in terms of both kind and amount, can be determined only by evaluating the individual on the job, much useful data can be gathered by observing as the content is presented. During the development of instructional materials, every effort should be made to avoid unnecessary duplication of content, gaps in content, and conflicts in concept. However, discrepancies are bound to occur, and these weaknesses must be identified and eliminated. Classroom observation is one means of doing this. The comments of instructors and trainees are another means of ensuring that the content is appropriate. In addition, the appropriateness and level of the content presented can also be judged by observing the trainees and the instructors as they work with the materials.

Sequence and Time Allocations

The only practical way to determine the correctness of the sequence of instruction and the amount of time allotted to each block of instruction is to subject these system ele-
ments to actual trial. Trainees will quickly note blatant errors in sequencing, because they will have difficulty with new material if they do not possess the knowledge and skills required to deal with it. Instructors will easily note cases of improper time allocation when they attempt to develop specific knowledge and skills. The comments of both trainees and instructors, as well as the observations of an outside evaluator, will provide the data needed to improve both sequencing and time allocations.

**Instructional Strategies**

The most reliable method of determining the optimum instructional strategy is to conduct experimental studies in which different strategies are compared. This is an expensive and time-consuming procedure. Nonetheless, some such studies should be conducted, but the number of controlled studies to be undertaken will be limited by personnel, time, and funds. For this reason, much of the initial data pertaining to the adequacy of particular instructional strategies must be obtained by observing the system in action. Along with observation, the judgments of all involved—trainees, instructors, and evaluators—must be recorded.

**Materials, Equipment, and Facilities**

Judgments about the adequacy of materials, equipment, and facilities can be made prior to the setup of the system. However, valid judgments about the appropriateness of these system components, in terms of both kind and amount, require observation of the system in action and the comments of those directly involved. Part of the job of an evaluator is to note deficiencies in these areas, collect the judgment of trainees and instructors, and make recommendations for improvement.

**Perspectives Required**

A training and development system can be evaluated from several different but complementary perspectives. Unfortunately, in the great majority of training programs, only one or two of these perspectives are represented. The most common is the perspective of the training manager, with that of the trainee running a close second. The training manager is not directly involved in the training, although he or she must assume responsibility for its success or failure. His or her view, then, is that of an outsider to the learning situation. The trainee perspective, although frequently a matter of concern, is often haphazardly surveyed. The means of collecting the data are not sufficiently well planned, typically lack objectivity, usually deal with only the broader elements, and therefore rarely provide information that is usable for improving the system.

Even if these two perspectives are adequately represented and the data collected are usable, additional perspectives are needed. These include the observations and evaluations of instructors, trained evaluators, and line supervisors.

**The Trainee Perspective**

Trainees are often in the best possible position to provide insights into the operation of a training system. Full advantage of this source of feedback should be taken by periodically administering and analyzing trainee questionnaire surveys and by conducting in-
Evaluating Training Systems

Interviews with trainees. Although it is true that trainee reactions, opinions, and judgments frequently lack objectivity and reliability, the trainee perspective must be represented. Careful development of the data-gathering devices and skillful use and intelligent interpretation of findings will go a long way toward improving the objectivity and reliability of data collected by this means.

The Instructor Perspective

Instructors, too, are an integral part of the learning system. For this reason, their observations and judgments are necessary inputs to an evaluation program. Furthermore, because of their training and experience, they can provide feedback to systems designers that is extremely valuable. Although bias is always a danger because of the instructors' desire to "look good," it is essential to collect and analyze the judgments of those who are—except for the trainees—closest to the learning situation. The degree to which instructors can render objective judgments depends on the area to be investigated and the means used to gather the data. Instructors can provide the best estimates of the adequacy of time allocations, sequence, strategy, equipment, and materials. Their judgments are equally reliable with regard to trainee input and the adequacy of facilities. Such information should be collected by having instructors complete rating scales and questionnaires, and by scheduling periodic interviews.

The Training Evaluator Perspective

One of the best sources of valid and reliable data about the effectiveness of system components is experienced training evaluators. They will be either instructors or instructor-supervisors and hence have the knowledge and skills essential to proper and complete evaluation. In addition, evaluators are able to be more objective about the evaluation because they are not directly involved in the training system. By means of observation, rating, and interviewing, evaluators collect data on the effectiveness and efficiency of various system components.

The Training Manager Perspective

The training manager is ultimately accountable for the effectiveness and efficiency of training and development systems designed, developed, and operated in the enterprise. Although he or she has this "vested interest" in the success of the systems, his or her education, training and experience in designing and conducting training and development programs permit professional appraisal of the effectiveness of systems components and their interaction. His or her special contribution must be exploited. The techniques used by the training manager include observation, rating, and interviewing.

The Line Supervisor Perspective

Rarely are line supervisors involved in internal evaluation of training or development programs. This is a mistake. Line supervisors have not only an interest in the kind of training their subordinates or potential subordinates receive but also expertise in the job or skill being taught. They can render invaluable assistance to training systems designers by observing and reporting the adequacy of objectives, content, equipment, and materials. Although line supervisors may not be qualified to comment on the strategies used
or the effectiveness of other components of the learning system, their contributions to appraisal are invaluable and should be obtained.

**Pitfalls in Evaluation**

Too often, programs of evaluation based on the use of observation and evaluative instruments have failed. These failures can for the most part be attributed to inadequate planning, lack of objectivity, evaluation errors of one sort or another, improper interpretation of findings, and inappropriate use of results.

**Poor Planning**

To be effective, a program of internal evaluation must be carefully planned. Some of the most common deficiencies in planning are:

- Failure to work out the details of the program, failure to include data-collection instruments, specific procedures to be followed, and the scheduling of observation, surveys, and interviews
- Failure to train evaluators in the principles and techniques of evaluation, including the use of data-gathering instruments
- Failure to make clear to all concerned the purposes of the evaluation program and the uses to be made of evaluations and recommendations

**Lack of Objectivity**

Although it is impossible to guarantee that evaluations will be completely objective, there are some steps that can be taken to make certain they will be more objective:

- Select evaluators who are capable of making objective judgments.
- Train evaluators.
- Design appropriate data-gathering instruments.
- Look at all the components of the learning situation as an integrated system.
- Focus on important details—avoid nit-picking.

**Evaluation Errors**

When scales are used to evaluate quality of performance, traits, or materials, observers often differ in their ratings. These differences are called evaluation errors, although this may not be the most accurate term to use for all these disparities. Some errors are caused by faults in the design of the rating instrument; others are caused by the raters. Some errors occur only with certain groups of observers; some occur only with individual observers. Other errors occur only when certain traits of individuals are rated. Some observers make errors when rating all individuals, some when rating certain groups, and others when rating certain individuals.

Rating errors can be classified into four general categories: central tendency, standards, halo, and logical.
Error of Central Tendency. Many evaluators are reluctant to assign either extremely high or extremely low ratings. Their ratings tend to cluster close to the middle of the scale. This error is most commonly made by inexperienced observers. However, even experienced raters can make this error when they rate personal qualities or abilities that are intangible, for example, leadership ability and teaching ability.

The tendency to give extreme ratings is opposite of central tendency, but it is considered to be in the same category. Occasionally, an observer will place too many ratings at the extremes of the scale. Everyone is identified as either superior (error of generosity) or unsatisfactory (error of parsimony).

Error of Standards. Some observers tend to overrate or underrate everyone in comparison to the ratings of other qualified judges. This is because their standards are either too high or too low. Experience with, and training in, the use of a particular rating scale usually results in similar distributions of scores by several raters, indicating that their standards are compatible.

When differences in standards are consistent and have enough stability to permit correction, the error is called systematic, or constant, error. Although this kind of error can be adjusted by adding or subtracting the same amount from all the scores of a given rater, a better solution is to provide further training to that rater so that he or she can correct the error.

Error of Halo. Some raters are unable to prevent a general impression they have of an individual from influencing their scoring of the individual's performance or traits. Usually this impression that clouds their judgment is from prior observation or knowledge of the individual being rated. However, likes, dislikes, and prejudices may also cause errors of halo. Raters' reactions to physical features, race, or nationality can influence their general impression. It is important to note that halo error can be favorable or unfavorable; therefore, it can result in either a higher or a lower score than the actual performance of the individual warrants.

It has been frequently observed that when people rate close friends, they tend to rate them higher than they should in all traits. This is called error of tendency. When halo error is traced to such sources as physical features, race, or nationality, it is called error of stereotype. Halo errors frequently go undetected; even when found, they are extremely difficult to correct.

Logical Error. This type of error, sometimes called error of ambiguity, occurs when two or more traits or abilities are being rated. If the rater sees certain traits as related (although the relationship may be obscure or illogical to other qualified raters), similar ratings for the different traits or abilities will be given. For example, some people believe that a person who is intelligent is also creative, or that a person who is industrious is also efficient. Intelligent people may be creative, but not necessarily. Similarly, industrious people may or may not be efficient. Usually the person who exhibits logical error is not aware of his or her fault.

Improper Interpretation

The collection of data is one thing; interpreting the data is quite another. Here the meaning and impact of the data are judged. If this step is not handled properly, the value of the information collected will be completely lost. Some of the main pitfalls in interpretation are:
Assuming that consensus among one category of observers on a single system element guarantees a valid and accurate judgment.

Concluding that an observation or judgment made by only one observer is inaccurate or invalid.

Taking comments at face value and not considering the nuances of language and the problems of semantics.

Failing to take into consideration the perspective of the individual making the observation.

*Inappropriate Use of Results*

When tabulated, data collected during internal evaluation have the aura of complete objectivity and truth. Sometimes the results of evaluation are used for purposes other than that originally intended. This is a major error. Some of the inappropriate uses to which evaluative data have been put are:

- Using ratings and reports of observation as the basis for disciplinary action.
- Using ratings and reports designed for systems evaluation as a basis for denying or granting special privileges or promotions.
- Using otherwise unsupported and unvalidated observations as a basis for making significant changes in an instructional system.

*Means of Evaluation*

*The Evaluation Process*

The first step in the evaluation process is to determine the specific aspects of the training and development system to be evaluated. Then the means and instruments to be used in collecting the data must be selected. After the data have been collected, they must be tabulated and summarized. Then they are both analyzed and interpreted, at which time recommendations are made and an improvement plan is drawn up and implemented. This section describes the most useful means of evaluation.

*Observation*

Observation is a method of determining the overt behavior of people as they act, interact, and express themselves in a situation selected to typify normal conditions. Observation is the most direct means of studying trainees, instructors, and the conditions that surround learning. It is the only way that certain aspects of a training system, notably the interaction of system components, can be studied.

Observation, in the context of training system evaluation, has certain characteristics:

- *It is specific.* Observation is not just looking around or seeking general impressions. To be useful, there must be carefully defined things to look for.

- *It is systematic.* Observation is not just dropping in on a training situation. The timing of observations, the length of the periods, and the number of observations must be carefully planned and scheduled.
Evaluating Training Systems

- It is quantitative. Insofar as is possible, measurable characteristics are the object of study in observation used for evaluation.
- It is recorded. A record is made of observation either during or immediately after the visit to the classroom or training area. The results and findings are not entrusted to memory.
- It is expert. Observation is conducted by fully qualified personnel who have been specially trained for the task.

During system validation, evaluators should observe as many classes as possible during each pilot course. Ideally, evaluators should be assigned so as to ensure 100 percent coverage. As an absolute minimum, 50 percent of all scheduled instruction should be observed and evaluated.

In planning for observation, evaluators should ensure that all critical points in the instructional sequence are observed and that an adequate sample of all other instruction is included. Critical points in the instruction are the times when trainees are introduced to important job tasks or skills and the times when they perform the job task or skill in a go, no-go situation. An adequate sample of other instruction is defined as observation of day-to-day instruction leading up to the critical performance. Figure 20-1 contains a sample observation report form that could be used to document and report these observations.

Personnel who perform observation should be fully qualified instructors, instructor-supervisors, training managers, or line supervisors. All must have received training in the techniques of observation and reporting.

Ratings

Periodically during each pilot course, various elements of the training system should be rated independently by several qualified raters. These elements include trainees, instructors, equipment, materials, training aids, and facilities. A sample scale for rating a course or block by trainees is shown in Figure 20-4. Scales such as these should be prepared before the pilot courses begin.

Personnel using the rating scales must have received thorough training in the principles and procedures of rating, including some training in how to avoid rating errors, and should have had supervised practice in the use of these scales.

Trainee Surveys

Trainee judgments are valuable in determining the effectiveness of certain training system components. Although they may be less objective than the evaluations of others involved in the system, they should be surveyed. However, such surveys must be used in combination with other evaluative techniques to ensure the reliability and validity of the judgments expressed.

At the conclusion of selected blocks of instruction and at the conclusion of the course, trainee questionnaire surveys should be administered to all members of the pilot classes. Instruments similar to the one shown in Figure 20-3 should be used for this purpose. Additional surveys may be administered as the need for them becomes apparent, for example, when it is necessary to study smaller segments of the system.

(Text continues on page 462.)
Figure 20-1. Sample observation report form.

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Students' aptitude for the instruction</td>
<td>Lacking</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>2. Students' readiness for the instruction, in terms of background</td>
<td>Lacking</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>and experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Students' motivation and interest</td>
<td>Lacking</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>4. Instructor's knowledge of subject matter</td>
<td>Inadequate</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>5. Instructor's teaching skills</td>
<td>Inadequate</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>6. Relationship between content and objectives</td>
<td>Inconsistent</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>7. Accuracy of content</td>
<td>Inaccurate</td>
<td>Questionable</td>
<td>Accurate</td>
</tr>
<tr>
<td>8. Level of instruction</td>
<td>Inappropriate</td>
<td>Questionable</td>
<td>Appropriate</td>
</tr>
<tr>
<td>9. Sequence of instruction</td>
<td>Inappropriate</td>
<td>Questionable</td>
<td>Appropriate</td>
</tr>
<tr>
<td>10. Time allocation</td>
<td>Improper</td>
<td>Questionable</td>
<td>Proper</td>
</tr>
<tr>
<td>11. Instructional strategy</td>
<td>Inappropriate</td>
<td>Questionable</td>
<td>Appropriate</td>
</tr>
<tr>
<td>12. Instructional materials</td>
<td>Unsuitable</td>
<td>Questionable</td>
<td>Suitable</td>
</tr>
<tr>
<td>13. Instructional equipment</td>
<td>Inadequate</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
<tr>
<td>14. Classroom facilities</td>
<td>Inadequate</td>
<td>Questionable</td>
<td>Adequate</td>
</tr>
</tbody>
</table>

Remarks

<table>
<thead>
<tr>
<th>Item number</th>
<th>Description of deficiency</th>
<th>Recommendation</th>
</tr>
</thead>
</table>

Evaluator
Figure 20-2. Sample scale for rating trainers in a course or block of instruction.

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Course or block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Location</td>
</tr>
</tbody>
</table>

Directions: Rate the course or block on each of the items listed. Place an X on each line at the place which seems to you to be the most appropriate. The highest possible rating for any item is 5, and the lowest is 1. To help you make your rating, three descriptions are given for each item. The one at the left identifies the highest rating, the one at the right identifies the lowest, and the one in the middle identifies the average rating.

**Objectives**

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal clear</td>
<td>Reasonably clear</td>
<td>Not at all clear</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Importance of course or block to job**

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critically important</td>
<td>Of average importance</td>
<td>Not important</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Course or block organization**

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceptionally well organized</td>
<td>Satisfactorily organized</td>
<td>Poorly organized</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sequence of topics**

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well sequenced, easy to follow</td>
<td>Satisfactorily sequenced</td>
<td>Poorly sequenced</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Course requirements**

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just right</td>
<td>Reasonable</td>
<td>Unreasonable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Methods and techniques

<table>
<thead>
<tr>
<th>Score</th>
<th>Varied and very effective</th>
<th>Occasionally inappropriate, more variety needed</th>
<th>One method used exclusively</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Grading

<table>
<thead>
<tr>
<th>Score</th>
<th>Fair and impartial, appropriate evidence</th>
<th>Partial sometimes, based on limited evidence</th>
<th>Partial, no real evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Instructor's knowledge of subject

<table>
<thead>
<tr>
<th>Score</th>
<th>Broad, accurate, up to date</th>
<th>Somewhat limited, not always up to date</th>
<th>Seriously deficient, out of date</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Level of trainee interest

<table>
<thead>
<tr>
<th>Score</th>
<th>Unusually high, completely attentive</th>
<th>Mildly interested and attentive</th>
<th>Disinterested, inattentive</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### General estimate of instructor

<table>
<thead>
<tr>
<th>Score</th>
<th>Superior</th>
<th>Average</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### General estimate of course or block

<table>
<thead>
<tr>
<th>Score</th>
<th>Very useful, effective</th>
<th>Average in usefulness, average in effectiveness</th>
<th>Extremely limited in usefulness, ineffective</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Other comments
Figure 20-3. Sample course or block evaluation questionnaire for trainees.

1. Did you get more or less out of the course or block than you expected?  
   If less, what was missing?
2. What specific topics in the course or block should have been  
   a. Omitted?  
   b. Added?  
   c. Emphasized more?  
   d. Emphasized less?
3. Which techniques of instruction (lectures, demonstrations, discussions,  
   practical exercises, case study, programmed instruction, etc.) did you get  
   a. The most from?  
   b. The least from?
4. Was the text helpful? If no, why not?

5. Were the handouts helpful? If no, why not?

6. What was your reaction to the tests and quizzes?

7. In what specific ways could the course or block be improved?

8. In what specific ways could the instructor improve his or her performance?

9. Would you recommend this course to your associates? If no, why not?

10. Please add any other comments you would like to make about any aspect  
    of the course or block.
Trainee Interviews

Personal interviews with trainees can provide information not available by any other means. With skillful questioning, trainees can be encouraged to express themselves fully and freely about the training program, their attitudes toward it, and its strengths and shortcomings. As with questionnaire surveys, interviews should not be used alone. The data they provide should be corroborated by other evaluative techniques.

During the progress of each pilot course, evaluators should conduct interviews with a sample of at least 10 percent of the class. Standard schedules should be designed prior to the start of the pilot course for this purpose.

Only personnel who have been thoroughly trained as interviewers should conduct these interviews. Personnel selected for this training should have demonstrated (1) a knowledge of training principles and practices, (2) an understanding of the system as designed, (3) an ability to conceptualize and analyze, (4) an ability to formulate questions, (5) an ability to gain participant response, and (6) a capacity to withstand stress. Supervised practice in conducting interviews is the only means of ensuring that personnel can meet these requirements. A nonstandardized interview form is presented in Figure 20-4.

Group Interviews

Focus groups, long used in marketing research, provide a quick, inexpensive, and flexible means of making qualitative judgments about the effectiveness of training when objective, quantitative data are lacking. The technique is particularly useful in such difficult-to-measure areas as management and supervisory training. Although focus groups are time-consuming, they usually bring out information and insights that are not revealed by other methods.

Essentially, a focus group is an interactive evaluation method designed to obtain in-depth answers to complex questions. Under the guidance of a leader or facilitator, the group discusses the training they have undergone and share their thoughts, ideas, attitudes, and opinions. The results are qualitative rather than quantitative.

Effective focus group evaluation has these characteristics:

- It is guided by a skilled and experienced leader or facilitator, preferably someone from outside to ensure candid feedback from participants.
- The process is standardized; the problem is defined, the questions are pre-planned, and the participants are carefully selected.
- It uses more than one group, with from six to twelve members in each group.
- It aims to generate frank and unbiased feedback.
- Comments of participants are recorded and analyzed and interpreted.
- It includes a review procedure and a written report of findings, conclusions, and recommendations.

Instructor Surveys

Instructors, too, have much to contribute to the evaluation of a system in operation. Their observations and recommendations should be systematically collected to ensure that the system is consistent with the needs of the implementers of the training.
Evaluating Training Systems

Figure 20-4. Sample nonstandardized interview form (trainees or instructors).

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>Instructional unit</td>
</tr>
<tr>
<td>Department</td>
<td>Division</td>
</tr>
</tbody>
</table>

Comments

Student input:

Instructors:

Course content:

Sequence and time allocations:

Instructional strategies:

Materials, equipment, and facilities:

Interviewer

Therefore, at the conclusion of each block of instruction and at the end of the course, questionnaires should be completed by all instructors. A standard questionnaire form similar to the one shown in Figure 20-5 can be used for this purpose.

Instructor Interviews

Although instructors are usually less reluctant than trainees to make written judgments about a training system, their involvement is so direct that they may be hesitant to put down their true opinions. For this reason it is extremely profitable to interview instructors in order to draw out their true feelings and obtain their judgments about each major component of the learning system.

Therefore, at the conclusion of each block of instruction and at the end of the course, at least a sample of instructors should be interviewed in depth. Personnel with the same qualifications and training as those used for trainee interviews should conduct these interviews.
Figure 20-5. Sample questionnaire for instructors.

1. Were you satisfied with the performance of the group? What specifically were the shortfalls, if any?

2. What specific topics in the course/block should be
   a. Omitted? ______________________
   b. Added? ______________________
   c. Emphasized more? ______________
   d. Emphasized less? ______________

3. In terms of time allocations, what specific lessons
   a. Require more time? ______________
   b. Require less time? ______________

4. What changes, if any, do you recommend be made in the sequence?

5. What changes do you recommend in instructional strategies?

6. What materials should be
   a. Added? ______________________
   b. Eliminated? __________________
   c. Revised? ___________________

7. What equipment should be
   a. Added? ______________________
   b. Eliminated? __________________

8. What changes do you recommend in the evaluation strategy?

9. In what other ways could this course or block be made more effective?
Evaluating Training Systems

Procedures for Collecting Data

The collection of data involves the application of a number of evaluative techniques and devices, including observation, rating, questionnaire, and interview. The most effective technique should be selected to collect the required information. Great skill is necessary if the data collected are to be pertinent, objective, and error-free.

Observation

Step 1. Develop a plan for observation; this plan should indicate the number of observations, their length, and their spacing. In scheduling observation, observe these rules:
   a. Include those activities that are truly typical of the system under study.
   b. Include critical instruction.
   c. Include enough day-to-day instruction to obtain a complete picture of the system in operation.

Step 2. Review the lesson plan and supporting materials before arriving at the classroom or the training area. Be sure that the objectives of the instruction are clear in your mind.

Step 3. Arrive at the classroom or training area before the instruction begins, and remain long enough to get a clear picture of the learning situation so that the objectives of the observation can be accomplished.

Step 4. Select your position in the training area carefully. Be sure that you can see and hear what is going on, but are not in the way.

Step 5. Make every effort to avoid being a distraction to either the instructor or the trainees. Obey these rules:
   a. Be as quiet and unobtrusive as possible.
   b. Refrain from commenting on content or procedures during the presentation.
   c. Try to convey an attitude of interest, attention, or objectivity.
   d. Guard against display of disagreement, displeasure, or boredom by facial expression, gestures, or posture.

Step 6. Focus your attention on each component of the learning situation, but pay particular attention to trainee reaction and performance.

Step 7. Complete the record of observation as soon as you leave the classroom or training area.

Rating

Step 1. Select the aspects of the training system to be rated.
Step 2. Establish a schedule for rating, with dates, times, and raters indicated.
Step 3. Whenever possible, use multiple ratings; that is, use several raters and average their ratings.
Step 4. Make certain that all raters understand the use of the scale. Points on the scale should have a common meaning for all users.
Step 5. Rate each person or item in comparison with all others in the group. If only one person or item is being rated, compare the person or the item mentally with others of the same level, type, class, occupation, or the like, that are not being rated.
Step 6. Rate each person or item on one trait before going on to the next.
Step 7. Use all steps in the rating scale, even the extremes if they are deserved.
Step 8. Rate only after you have had enough time to observe the individual's performance.
Step 9. Do not rate individuals or traits for which you cannot cite specific evidence to support your rating. If you have no basis for judgment, do not rate that item. Mark it "No opportunity to observe."

Questionnaires

Step 1. Determine the purpose of the questionnaire.
Step 2. Develop a survey plan and schedule; include the who, when, and how.
Step 3. Select respondents by name.
Step 4. Distribute the questionnaires.
Step 5. Follow up if returns are not received within a reasonable time.
Step 6. Process and interpret the data.
Step 7. Compare the data with data from other sources.
Step 8. Prepare a summary report. (Even if respondents have been asked to sign the questionnaires, the respondents' names are never included in the summary report.)

Interview

Step 1. Determine the objectives of the interview; know specifically what it is that you are trying to find out.
Step 2. Select the interviewees by name.
Step 3. Learn as much as you can about each respondent in advance; study records, reports, test results, as well as reports of earlier interviews, if there are any.
Step 4. Set up an appointment with each interviewee at a mutually convenient time.
Step 5. Select a place for the interview that is comfortable, private, and free from the distractions of noise and interruptions.
Step 6. Conduct the interview. When interviewing, observe the following rules:
   a. Explain the purpose of the interview and the use to which the findings will be put.
   b. Establish a friendly, cooperative working relationship with the respondent before getting to the task at hand; put the interviewee at ease. Emphasize that the interviewees' identities will be protected and that all their comments will be anonymous.
   c. Begin with questions that are easy to answer and are not emotionally loaded.
   d. Let the interviewee talk; let the interviewee feel free to express himself or herself; do not dominate.
   e. Deal with the interviewee in a forthright, sincere manner; do not be pedantic, and do not try to be shrewd or clever.
   f. Avoid evidences of pressure, boredom, and irritation.
   g. Give the interviewee an opportunity to qualify his or her answers.
   h. Ask only one question at a time.
   i. Avoid antagonizing, embarrassing, or hurrying the interviewee.
Evaluating Training Systems

j. Do not push ahead too rapidly, but do not dawdle.
k. Keep control; do not allow the interviewee to go off on extended tangents.
l. Display an objective but not an uninterested attitude.
m. Be alert for leads; watch facial expression, gestures, and casual remarks.
n. Do not be perturbed by expressions of negative feelings, such as hostility or highly subjective criticism.
o. Encourage the interviewee to state his or her views completely without fear of censure or reprisal for honest criticism.
p. Raise questions to elicit responses about areas not covered in the interviewee’s responses.
q. When the interview is over, summarize the main points to be certain that you have the facts as the interviewee presented them.
r. Record all data immediately.

Focus Group Interviews

Step 1. Skilled and experienced facilitators from outside are selected to lead the groups.
Step 2. Participants in each group, numbering from six to twelve, are carefully selected.
Step 3. Standardized questions are framed in advance of the meeting of the groups.
Step 4. The facilitator establishes rapport with the group and encourages all participants to provide objective feedback.
Step 5. The comments of participants are recorded, analyzed, and interpreted.
Step 6. A written report of findings, conclusions, and recommendations is prepared, distributed, and used to improve training.

Tabulating, Interpreting, and Using Evaluative Data

Tabulating and Summarizing Data

The problems of tabulating and summarizing the results of observations, ratings, questionnaires, and interviews are numerous and varied. These difficulties apply whether the tabulating is done manually or by machine. Where the data are quantitative, the problems are relatively simple. However, some of the data collected for internal system evaluation are nonquantitative description. With these kinds of data the selection of appropriate summarizing categories is much more difficult. In any event, tabulating and summarizing data are slow and demanding tasks. Accuracy is the key word.

The steps to be followed in tabulating and summarizing data are these:

1. Edit the completed forms and scales. Be sure that the responses on different parts of the same form or return are consistent and that they do not controvert known facts.
2. Establish summarizing categories for both closed- and open-end responses.
3. Determine the mathematical treatment to be applied to each summarizing category, for example, mean, median, mode, percentage, range, rank order, percentile, or standard deviation.
4. Make a preliminary list of responses under each summarizing category for the open-response type of item, in order to determine the nature and range of responses.
5. Establish a final list of responses for each summarizing category.
6. Study each response on all returns (item 1 on all forms, then item 2, item 3, and so forth), and tabulate them in the proper category by placing a tally beside the item.
7. Sum the tally marks in each category for each item and/or apply the mathematical function required.
8. Prepare a summary report of significant items.

**Interpreting Data**

The interpretation of data obtained from reports of observations, ratings, questionnaires, and interviews is not a simple, routine procedure. On the contrary, it is the most demanding and the most crucial step in the whole evaluation process. In fact, in the strictest sense, interpretation of the data is evaluation.

This, then, is the stage that depends most heavily on the professional knowledge and the skills of the evaluator. If the data are complete and accurate, the skillful evaluator will be able to identify significant weaknesses (if any) and draw conclusions on which recommendations for improvement can be based.

Because of the almost infinite variety of ways of expressing the data, it is impossible to describe all the difficulties that may be encountered in the interpretation phase. An example must suffice.

The interpretation of frequency counts is a common requirement, and it presents immediate and formidable problems. For example, if 75 or even 80 percent of a group of trainees state that they believe a certain block of instruction should be eliminated, what does it mean? Are the trainees right or wrong? What should the evaluator conclude? On what basis should a recommendation for the retention or elimination of the block be made to the systems designers?

Although frequency of mention may very well be an element of importance in reaching a conclusion, it is only one element, and its significance must be evaluated by carefully analyzing its logical and functional contribution to the total picture.

Evaluative instruments are based on two fundamental assumptions. First, it is assumed that more objective judgments can be obtained about the significant aspects of a training or development system by focusing on one component or element at a time. The second assumption relates to instruments that yield a total or composite rating. It is usually assumed that total value can be calculated by summing the values assigned to the individual parts. As for the first assumption, the question remains: Are the items selected for evaluation the really critical ones? It is entirely possible that in designing the instruments, certain crucial items were overlooked. As for the second assumption, even if all significant elements have been included, can any mathematical function applied to the parts yield a meaningful total value?

For these reasons and others, evaluation based on the results of using a combination of appraisal instruments is likely to be more accurate, valid, and reliable than evaluation based on the use of any single instrument. In interpreting data, these steps should be followed:
1. Analyze the summaries for each item in turn, and write a brief statement of their meaning and possible significance.
2. Compare the statements derived from step 1 with the findings of other evaluative instruments dealing with the same category to determine areas of agreement and disagreement. Consider the nuances of language, the problem of semantics, and the perspectives of the respondents.
3. Draw your conclusions, and state them in simple and concise language.
4. Draft your recommendations.

Using Results

The ultimate purpose of internal evaluation must be kept firmly in mind when the final product is to be put to use. Note that the purpose is to improve the training or development system as a whole by improving its components. The value of evaluation, then, depends on what is done with the conclusions and recommendations that the evaluation system produces.

The steps to be followed in using the results of evaluation are these:

1. Submit recommendations for system changes to the systems designers, the training manager, and the instructors for study.
2. Schedule a meeting of the functionaries identified in step 1 to discuss the findings and recommendations.
3. Establish a priority listing of changes.
4. Determine the resources required to implement the changes.
5. Submit the recommendations, priority listing, and required resources to the training manager for approval.
6. Establish who should institute the changes, when they should be instituted, and how they will be accomplished.
7. Acquire the needed resources, and institute the changes.
8. Follow up and re-evaluate the changed components of the system.

CHECKLIST FOR EVALUATING TRAINING SYSTEMS

Evaluation Components and Perspectives

- Are all the following components of the training system evaluated?
  - Trainees?
  - Instructors?
  - Course content?
  - Sequence and time allocations?
  - Instructional strategies?
  - Materials, equipment, and facilities?

- Are the following perspectives represented in training system evaluation?
  - Trainees?
  - Instructors?
  - Training evaluators?
  - Training manager?
  - Line managers and supervisors?
Evaluation Procedures

- Are the following pitfalls in evaluation avoided?
  - Poor or incomplete planning?
  - Lack of objectivity?
  - Evaluation errors?
  - Improper interpretation of data?
  - Inappropriate use of results?

- Are all the following means of evaluation employed?
  - Observation?
  - Ratings?
  - Trainee surveys?
  - Trainee interviews?
  - Instructor surveys?
  - Instructor interviews?

- Are procedures for tabulating, summarizing, interpreting, and using evaluative data clearly defined?

For Further Reading, Viewing, and Listening


Evaluating Training Systems


Kirkpatrick, Donald L. How to Evaluate Supervisory/Management Courses (120-min. videocassette), 1990. Distributed by American Society for Training and Development, 1640 King St., Box 1443, Alexandria, Va. 22313-2043.


Robinson, Dana Gaines, and Dennis C. Sweeney. Reaction Evaluations: How to Go Beyond the 'Smile Sheet' Response (90-min. audiocassette, 90AST-S8), 1990. Distributed by American Society for Training and Development, 1640 King St., Box 1443, Alexandria, Va. 22313-2043.


