

# THE EVALUATION OF TRAINING: A CHECKLIST APPROACH<sup>1</sup>

Michael Scriven  
Claremont Graduate University  
& Western Michigan University

Of everything that has ever been written about the professional evaluation of training, it seems to many observers, myself included, that Donald Kirkpatrick's 1959 contribution of the 'four levels' has lasted longest and deserved it best.<sup>2</sup> However, we have made some strides since then in developing the general discipline of evaluation, and I here propose some ways to elaborate his approach, based on those developments.<sup>3</sup> In the spirit of concision that he exemplified, I have tried to provide easily remembered labels for each of the bullet points of essential components given here. But this 12-point checklist as a whole, since it includes my substantial development—sometimes it's readable as a radical transformation that he might not have approved—of his four components (these are asterisked), makes no claim to concision. The eight I have added—some inspired by his—are fairly easy to understand and are accompanied by a few lines or paragraphs of explanation, examples of use and misuse, and something about methods of testing.

This checklist is intended for use when more than a simple bottom line is required, i.e., for what is sometimes called analytic or diagnostic evaluation, in either (i) formative or (ii)

---

<sup>1</sup> Acknowledgments. (i) This work was developed as part of a contract for external evaluation of the international work of Heifer Corporation on poverty reduction, when it became clear that Heifer's extensive efforts at training recipients of their donated livestock could benefit from improved evaluation. (ii) The checklist developed here has been substantially improved from its first draft form in response to some invaluable suggestions from Robert Brinkerhoff, to whom many thanks. Some of the points he made, and some others I borrowed from him, are to be found in his excellent book, *Telling Training's Story* (Berrett-Koehler, 2006). (iii) Some welcomed reactions by the ~60 members of EVALTALK who requested copies of this paper have also had an effect.

<sup>2</sup> The latest version of this is in his *Evaluating Training Programs: The Four Levels* (3rd Edition), Berrett-Koehler, 2006.

<sup>3</sup> Some of these developments are outlined in my *Evaluation Thesaurus* (Sage, 1991).

summative or (iii) ascriptive<sup>4</sup> mode. This approach can also be used for (iv) monitoring, where its use will often help head off serious problems. It may also be useful for (v) writing requests (RFPs) for evaluation of training, and for (vi) evaluating reports that are supposed to constitute serious evaluations of training (i.e., for meta-evaluation); and for (vii) helping in the design of good training programs. But it is particularly aimed at finding, or planning to avoid, situations where a training program might fail or has failed. With careful thought, one can also use the TEC to guide some educational evaluations, keeping in mind that education is different from training. How different? That's a very hard question; please see the endnote on the difference between training and teaching for one answer.

The checklist checkpoints are: 1. Need; 2. Design; 3. Recruitment; 4. Delivery; 5\*. Reaction; 6\*. Learning\*; 7\*. Retention; 8. Application; 9. Extension; 10. Value; 11. Alternatives; 12\*. Return on Investment (ROI). The full content is very detailed and in many contexts it would be too much to try for serious investigation of every checkpoint. But—an important 'but'—I believe it's worth thinking through each of them, to make sure you have at least general evidence that it's not going to be fatal to skip it in your detailed study. So do not give up when reading this because it seems excessively long and detailed for your purposes; mental checks cost almost nothing and often produce large and tangible returns.

I use the title 'Training Evaluation Checklist' (TEC), for this instrument, and, as mentioned, provide not only a definition of each checkpoint but also some indication of an appropriate method for verification, and often some common traps to be avoided. A forthcoming appendix (a.k.a. annex) to the paper will provide a hypothetical example of typical use in international aid training, covering two levels of cost. The low-cost version is intended to show that using the TEC need not be burdensome in straightforward cases, where reflection on many of the checkpoints is enough to tell you how to cover them without expensive investigation, and this is almost always cost-effective. The more elaborate kind of design is often well-justified when a large-scale or costly training project is involved (or proposed),

---

<sup>4</sup> Ascriptive mode is the use of evaluation simply to add to our knowledge. Examples include evaluations by historians of the way Napoleon used his cavalry, or road tests of the many cars that now cost more than a penthouse on 5<sup>th</sup> Avenue in New York (since few or none of their readers are candidates for purchase or redesign of such cars).

as it will often prevent common failures to deliver good results from training, or expedite their elimination when they do occur.

## TRAINING EVALUATION CHECKLIST (TEC)

Each of the following checkpoints should be addressed, even if only briefly, (i) in any serious *evaluation of*, or (ii) *proposal/design for*, a training program; and (iii) when *specifying* one to be developed or delivered by one's own organization or by others working for it; and when (iv) *reviewing an evaluation of* one.

**1. Need.** Here we look for *serious evidence* that training could really be the *best* answer to a *real* problem in *this* organization *right now* for *this group*. Requiring this kind of evidence seems obvious enough, but all too often the supposed need is merely one of the following: a long-established offering, not recently reconsidered; an unsubstantiated intuition by some executive who has been inspired by something they heard or read; the results of a *wants* survey<sup>5</sup> of staff (who may really be voting for a fun change from boring routines); something the HR department thinks would make them look with-it or at least useful; or a 'keep up with the Jones' response to some current fashion in training or anecdotal report from (or about) a competitor. 'Serious evidence' means being able to describe: (i) a *specified* increase in KSAV<sup>6</sup> that is (ii) *necessary* i.e., essential for survival, or at least highly desirable

---

<sup>5</sup> A *need* is something without which a system or person falters or fails to perform their function at an appropriate level. Needs are often not recognized as such by those who have them, whereas wants are felt preferences. Dietary needs vs. wants, especially in children, provide good examples of the difference. Staff surveys only provide a *felt needs* assessment for training. This will yield you some examples of true needs, but will miss key elements that you can only get from other sources, notably (i) supervisory ratings of performance competence, (ii) client/customer/consumer identification of weak performance, and (iii) objective data on performance errors or shortcomings.

Worse, it will include many responses that do not represent needs you should address, either because the respondents could not use the training in their jobs or could not understand the training.

<sup>6</sup> KSAV (pronounced 'kay-sav') is an extension of the usual shorthand KSA ((a.k.a. SKA), standing for knowledge, skills, and abilities), to include a V for values, including attitudes. (KSA is occasionally used to stand for knowledge, skills, and attitudes, which would cover values, but to avoid confusion, I have explicitly added values/attitudes.) Changes in values are sometimes trainable, at least to some extent—albeit with considerable difficulty—and often highly desirable, even life-saving. While there is some reluctance in the training profession to be up front about the need to change attitudes/values, it is obvious that this is in fact crucially important e.g., in safety training,

for optimal performance and is (iii) provably *trainable* for (iv) a *specified* group of people. Hence it should include test data, observations about performance, or other cogent arguments indicating that: (i) this group *does not now have* the requisite KSAV; (ii) this group is *capable of acquiring* the desired KSAV through training that we can provide or obtain; (iii) the training to do this would be *cost- and resource-feasible*; (iv) if the requisite training were provided, it would probably produce a performance improvement with *payoffs that would probably compensate for the projected costs*—meaning direct, indirect, and opportunity costs—of the training; (v) training by you or your chosen trainers is a *more cost-effective path to the desired state than at least the obvious alternatives* such as: (a) *hiring* new staff (for onsite or online work) who already have the required KSAV; (b) *outsourcing* the work to private or public educational/training organizations; and (c) providing more sophisticated *equipment* (e.g., computer hardware and/or software) for the present staff.

It is unusual to suggest that a needs assessment (a.k.a. needs analysis) should include considerations of feasibility and projected cost-effectiveness, but in most contexts (e.g., planning, monitoring, evaluating) there isn't much point in saying that someone 'needs' something that isn't feasible or that wouldn't be worth what it—or some cheaper alternative—would cost. Only if the context is one of fund-raising goals formulation, or of Gates-level resources, can one virtually ignore cost-benefit considerations. So requirements (iii), (iv), and (v) are usually appropriate.

It is also crucial to watch for and clarify situations where the payoff from the training will only occur if other groups, e.g., the trainees' managers, or 'reports' (a.k.a. subordinates), or peers, or customers/clients will support or cooperate in certain ways with *both* the trainees' release for training and with their new KSAV when trained. If either will be essential—and both are usually essential—one must ensure that they will in fact occur as needed, and provide evidence that this insurance has been obtained: or add them to the groups needing

---

anger management, sexual/racial/cultural harassment avoidance, addiction termination, entrepreneurship training. The hard parts are: (i) to correctly distinguish the values that can be legitimately targeted from others that are private rights; (ii) to make clear that, and why, value shifting is one of the aims (under informed consent constraints); and (iii) to produce or demonstrate any significant changes. Moral education is a supremely important example of training that involves all these difficulties.

training, probably of a different kind. (This will of course mean taking account of the consequent cost increase). This task should be regarded as part of the duties of the training/education/development department, if there is one—or the training consultant if not—and in the latter case will require some serious work by an appointed liaison staffer in the organization who has enough influence to get the required cooperation from departments in the organization.

Providing this kind of needs assessment in detail could be a major task requiring considerable skill, but it is typically much cheaper than undertaking training based on someone's hunches about these issues.

**The practical-level suggestion here** is only that each of the questions listed above, and in the following checkpoints, should at least be *addressed seriously in a management or staff discussion* (possibly involving a consultant) before a training proposal is requested.<sup>7</sup> Depending on the cost of the proposed or existing training, it may or may not be worth getting the needs assessment done professionally, either internally or externally, preferably using this checklist or an improvement on it.

**2. Design.** Under this heading, there needs to be not only a reasonably detailed design (i.e., one that specifies curriculum, pedagogical approach, staff KSAVs, and time/space/equipment requirements) but some evidence that the proposed training *and associated advertising, staffing, content, and required support*—is accurately targeted on: (i) the demonstrated need; (ii) the identified target group's background and current KSAV; and (iii) the resources available at the planned delivery site, including management and logistical support at all relevant levels. A prima facie test of the design can be done by carefully comparing the results from the needs assessment of Checkpoint 1 with the description and details of the proposed training plan, including the advertising, trainers, trainees, other staff required, and site. It is not enough to simply pick a well-qualified designer and assume s/he will produce a good training program, since good designers are often overcommitted and delegate such tasks to new and less competent staff, or fail to get site or trainee details, or cannot by themselves obtain the needed support from the rest of your organization. In

---

<sup>7</sup> Some of the preceding questions are addressed in more detail under other checkpoints below.

short, *you cannot assume that effective training will result from just hiring (or assigning) a good trainer.* Someone has to handle all the logistical details associated with the training, e.g., announcing, recruiting, ensuring attendance (including following up on non-attendance), site prep (including laptop outlets), coffee/drink/meals/provisioning,<sup>8</sup> projector/replacement bulbs/notebook/technician availability, etc.—and other support as described above and below. To keep down the frequency and scale of the most probable failures, make sure the contract or your own arrangements go beyond minima, and cover: (i) scheduling either a videotaping replay or a duplicate presentation of at least the first session *and* (ii) some kind of coaching or other support (e.g., an online or on-call expert) to follow up on particularly the first but also subsequent sessions of the training with assistance in implementation and other trouble-shooting; plus (iii) overkill-seeming proactive stimulation before the first session to get acceptable levels of attendance, participation, and implementation. (In other words, make sure you're not just providing or evaluating *training*, but a *system effort to make a change* that involves training.)<sup>9</sup> It is essential that coverage of all these significant matters is spelled out and assigned to an identified responsible manager and perhaps others.

**3. Recruitment.** While this of course includes announcements at staff meetings and in relevant inhouse newsletters and online sites, it involves a great deal more—and in one sense less. The key is to be sure to recruit all the people that most need the training but *not to get more*—because those who get the training but didn't need it will incur high opportunity costs at their normal worksites, and low returns in terms of benefits, *and* some of them will impede the efficiency of the training for others present. That's more of a recruiting effort than is commonly achieved—it will probably require serious efforts to give presentations at department meetings, not just emails to everyone—but there's still more. The announcements or recruiting presentations must not be misleading e.g., by overpromising what skills will be acquired, or for whom they will be provided, and must not be any longer

---

<sup>8</sup> The devil is in the details here; the most frequent gaffe used to be providing only sugar-laden accompaniments to coffee for groups that are certain to include many diabetics; the latest is to provide only vegan refreshments to groups certain to include unconverted and irritable carnivores.

<sup>9</sup> A recurrent theme in Robert Brinkerhoff's catechism on this subject.

than absolutely necessary because of the opportunity costs in work-time and/or negative effects on the image of the training staff for future efforts. Furthermore, refusing keen but inappropriate volunteers must be done diplomatically, sometimes not an easy task. There's more: good recruiting is open-ended and does not terminate when the right people for this training effort show up at the first meeting. It must be sustained by follow-ups to encourage further meetings or practice of what's learnt, and to deal with problems that arise with the materials, their application, and perhaps the reception of the training by fellow-workers and supervisors. The recruiter (who may or may not also be the trainer) must respond fast if attendance falls off, by tracking down the reasons via interviewing non-attendees, and taking remedial action as appropriate, including—if necessary—some changes in the training or trainer. At worst, if the problem is not remediable for this training, a lesson must be formalized for future practice.

**4. Delivery.** Here we need evidence that the actual training was in fact *announced, attended, supported, and presented as proposed* and/or promised in the description used to get the approval, funding, or contract (and perhaps also used to recruit participants, which gives it quasi-contractual status with respect to them). This needs to be checked by carefully comparing the contracted syllabus with: (i) the attendance record sheets, and (ii) the delivered preparation and contents as demonstrated by an audio- or video-tape plus the recipients' feedback from Checkpoint 5, or, preferably, by the personal observation of a *skilled* observer, preferably a participant observer. Proof of proper preparation and delivery should be a condition of at least half of the payment for the responsible contractor, who may or may not be the trainer(s). This is a good moment to mention that you need to have arrangements in the contract for settling disputes.

This is also a good moment to remark that a good observer should always be looking beyond the training being observed, since that process is often an excellent opportunity to identify trainees with special talents (or problems) of one kind or another, e.g., with leadership skills, or unusual inventiveness or resilience, or willingness to help fellow learners.

**5. Reaction.\*** Here we need evidence as to how the training *and* peripheral support was rated for relevance, comprehensibility, comprehensiveness, logistics, overall value, appro-

priate use of time, etc., by participants. Checking this should be done in the first place by using a well-designed and previously tested form that provides both closed- and open-ended questions, requiring no more than about 5-7 minutes to answer briefly (though it's preferable to allow 10-12 minutes in order to provide an opportunity for longer answers to open-ended questions). It's essential, although more difficult than most form-designers realize, to avoid bias in the way the questions are presented. (And no more smiley-face!) Although there is a point of view from which these responses are irrelevant *if* one is gathering evidence of 'real impact' (covered in later checkpoints), they are often an invaluable guide to identifying exactly what was problematic, and an early warning indicator of defects that will only show up much later in the evaluation via long-term outcome measurements, if you are able to get those at all. Moreover, getting immediate reactions is an often-appreciated sign of respect for the opinions of staff. Indeed, a conscientious effort, one that includes follow-up phone calls to ask for delayed reactions, is almost always worthwhile, i.e., it usually turns up matters needing—and repaying—attention. However, this checkpoint is sometimes treated as much too important—for example, it often provides the only evaluation data that is gathered at all, which is simply absurd. If training is evaluated like an entertainment item, you get shows without substance, and you deserve them.

**6. Learning.\*** Here we need evidence that participants in fact mastered (at least much of) the intended content, and acquired the intended value or attitude modifications. This should be checked in the first place by a well-designed mastery learning test at the conclusion of training.<sup>10</sup> *Here is one point where we must also pick up unintended as well as intended effects.* For this we will need the cooperation of the observer of Checkpoint 3. Identifying and verifying unintended consequences is likely to require some interviewing of par-

---

<sup>10</sup> Technical Notes. A. Do not summarize these results in terms of average learning: show the full distribution, since even if only 10% of the group learn 10% of the target KSAV, this may more than pay for the total cost of the training (i.e., avoid Brinkerhoff's 'Tyranny of the Mean' fallacy).

B. This test should or should not use matrix-sampling from a comprehensive item pool, depending on whether it is primarily important to record group achievement levels or also individual achievements. Since its use greatly reduces cost and time requirements, matrix sampling should be the normal approach, because in the evaluation of training we are not normally required to be doing (trainee) personnel evaluation as well, which is (almost) the only justification for not using matrix-sampling.



ticipants as well as skilled observation of process and at least one question in the questionnaire required by Checkpoints 5 and 7. If a videotape or audiotape is used, it's better if it is stereo with one microphone and channel dedicated to audience input. The optimal design here uses two observers for the task, one of them not informed of the intended consequences of the training (i.e., operating in 'goal-free mode'), reporting only on what s/he sees as occurring or apparently occurring. What was actually learned may be very different from what was taught, and this checkpoint *must* cover the former, not just the latter: this requires considerable skill, but is essential for serious evaluation. For example, participants may have learned how to make it appear that they have learned what was intended without actually mastering it. They may also have formed acquaintanceships—or even networks—of substantial later value; or formed impressions, accurate or not, as to what the organization's less explicit values are. Some of these possibilities should probably be covered by specific questions in the later revisions of the participant rating form and tests.

Note that in connection with this and the next two checkpoints particularly, it's important to read the Endnote, which discusses the two types of learning that are involved in training.

**7. Retention.** Here we must determine whether the participants retain their learning of knowledge, skills, attitudes, or values for an appropriate interval or intervals. (Sometimes they increase it!) For content where application is needed immediately, a follow-up test at 15 or 30 days might be appropriate; where long-term retention is important, 90 or 180 days, or up to 2 years, might be more appropriate, or at least included. In some cases more than one test or set of interviews may be desirable; in all cases, as with the Learning checkpoint, attention should be paid to finding unintended and/or unexpected consequences. Note that this checkpoint is not duplicating the next or previous checkpoint, which should supplant it only if all three cannot be done. If it's hard to do all three, try very hard to do them all in at least the first round of testing new training, to enable more accurate diagnosis of points of failure/success.

**8. Application.\*** Here we find out whether participants *appropriately used*, and *continued to use appropriately*, what they learned from the training in their work context (whether or not it was the intended learning). As with Checkpoint 6, this will need to be checked at an

appropriate interval after the training is concluded, but checking is of a very different kind from that required for 6. It will involve one or, very much preferably, more than one of the following: (i) observation of work performance; (ii) examination of work product; (iii) interview of supervisor; (iv) interviews of co-workers. In each case, the exact nature of the check may need to be quite sophisticated, and will need to be standardized after some trials. Note the very important family of cases where the training is capacity-expanding but the capacity is not intended for regular use—e.g., CPR training, physical disaster training (fire/earthquake/flood/attack), use of firearms to immobilize but not kill. These are cases where we want applicability, but we don't want frequent use of it, i.e., frequent application. For these we mainly rely on the test of retention in the previous checkpoint, not the observation of regular use in this checkpoint, *but* (i) we need to make sure those tests are highly realistic i.e., are almost always include simulations rather than just being paper-and-pencil tests; *and* (ii) if there *was* any application, we need very careful checking on the responses, for quality and quantity. Hence this checkpoint is always something to be addressed seriously and it doesn't render checkpoint 7 redundant. And without 7 you can't *locate* some failures, which you need for formative evaluation and most monitoring and recommendation purposes.

**9. Extension (a.k.a. Generalizability).** We should now add another kind of perspective—a look at the possibility that this training package can be usefully replicated for other training purposes in this or other contexts. By now, you will understand that the 'package' is a complex entity and refining its construction and delivery is a massive achievement. Its exportability is therefore a major issue in evaluating its significance/value. Extension here means, for example, its deliverability: (i) at other times (this may seem trivial, but think carefully about weather/religious holidays/deadline times etc.), (ii) other sites, (iii) using different staff as trainees or as trainers, (iv) in other organizations, or (v) with other subject matter. This is the *potential* payoff, by contrast with the *immediate* payoff. There are times when this consideration will in fact provide by far the most important benefit of the whole exercise, so it is worth serious thought—and it takes serious thought, the more so because at first it seems irrelevant.

**10. Value.** Here we need to consider the specific qualitative value of each component element of the impact of this training, particularly of those that were unintended. We estimate this value by integrating the magnitude and directions of each effect with its relevant values for the organization, the trainees, and the environment (social and bio-physical), taking into account some estimate of the importance of the value. This requires some of the special skills of an evaluator in the identification and weighting of values, which can be done either qualitatively and sometimes—if it's possible without distortion—quantitatively, and their integration with empirical results. The result of this analysis, at this point, will be a list of evaluative pros and cons of the training, with some indication of importance. Note here that there is a very important category of cases where the training is legally required or legally crucial for defense against possible suits for lack of due diligence, so providing it is virtually necessary, regardless of any probable economic or environmental payoff to your organization. (After all, you don't pay for insurance because it *does* pay off, only because it *might*.) We could call this 'insurance value;' it is an important part of sustainability (often with the incidental payoff of stress reduction for staff, and almost always for managers and sponsors).

**11. Alternatives.** A thorough evaluation now requires that the impact of the training, as just determined, be compared with the (measured or estimated) impact of known alternative approaches to meeting (more or less) the same needs that the training addressed. These might be other approaches to the same training, or ad hoc hiring, adding technology, or changing the duty allocations for existing jobs. This gives us an important perspective on the training at which we're looking, even if only rough estimates of the performance of the alternatives are possible.

**12. Return on Investment (ROI).** Finally we come to the overall return on investment for the organization, but calculated in five specified dimensions. These include a double extension of the so-called 'triple bottom line' approach, hence this approach might be called the 'quintuple bottom line.' The usual 'triple bottom line' is often expressed in the phrase "People, Planet, Profit"—which we might call the triple P version of the triple bottom line. I prefer the triple E version, which is not quite the same, meaning: (i) the Economic, (ii) the En-

vironmental (biological *and* social<sup>11</sup>), and (iii) the Ethical (*and* legal) dimensions of pay-off.<sup>12</sup> The additional two dimensions in my 5D version fit better into the triple E version. They are: (iv) the value of potential Extensions of the approach to other contexts or uses (from the Extension checkpoint); and (v) the comparison of this approach with the alternative possible approaches to the same ends (from the Alternatives checkpoint);<sup>13</sup> this is the approach's 'comparative value' or, to keep the alliteration intact, the Exclusiveness value. Hence: Economic/Environmental/Ethical/Extension/Exclusiveness.

In this last checkpoint we pull together and try to integrate *all* the costs and benefits, quantitative (where possible) and qualitative (at least), on *all* of these dimensions, and get the total net impact of the training program, from the point of view of the organization—but with due regard to the impact on others and the environment—which may be positive or negative. Attractive though a completely quantitative approach certainly is, the reality is that only a hybrid will be possible in the vast majority of cases, especially with respect to the ethical dimension. But that does not exclude clear and provable answers: as we will see in the example that follows, such answers are often achievable, and can usually be rendered graphically even if not numerically. At least sometimes, a picture is worth not only a thousand words, but a thousand numbers.

One other element needs to be included in our work before we sign off on the complete evaluation, and it can be thought of as part of the process of dealing with the wrap-up checkpoint, Checkpoint 12, or as a separate Checkpoint 13, the lucky or unlucky number. This is an external review of the evaluation itself, in other words, a meta-evaluation—and any improvements that result from it. With any major evaluation, and with lesser ones if possible, a small expenditure on getting a skilled evaluator (of training) to critique one's

---

<sup>11</sup> Note that in calculating social impact, changes in human and social capital must be included; and note that in all dimensions, sustainability must be considered very carefully.

<sup>12</sup> An excellent balanced account of the *triple bottom line* approach is in Wikipedia (at 8/08). The best-known enthusiast account is probably *The Triple Bottom Line* by Andrew Savitz (Jossey-Bass, 2006). My separation of the ethical/legal dimension is novel, but it seems reasonable.

<sup>13</sup> Why does this get into the ROI? Because you want to find out whether the investment in this training program provided something the client organization could not have got in another way, cheaper, better, faster, or more easily. Establishing the merit and value of the program requires that.

own evaluation—usually just at the level of a half day or one day assignment—is one of the best investments one can make in improving the evaluation of training. This is especially true when setting up an evaluation approach that will be used on more than one occasion, as occurs when a new training program—or a new approach to evaluation—is started up. (Robert Brinkerhoff was kind enough to perform that function for the effort here.) My hope is that the TEC will serve as a useful instrument for doing meta-evaluation in the future, and I hope that suggestions for improving it will be sent to the author by those who use it, and also by those who simply read it critically.<sup>14</sup>

**Example.** Heifer’s primary intervention consists in donating pregnant livestock to members of a group of marginal rural farmers, who have some experience with, and enough land to raise, the farm animals in question. These farmers have agreed to work as a group and to pass on, to a family in similar circumstances, the same number and quality of livestock, from the offspring of the gift they receive. In the interests of sustainability, Heifer always provides them with some training in livestock care and management and in a set of ethical and environmental values called the Twelve Cornerstones. The question has arisen whether they should also provide training in other self-help and group survival skills such as money management, micro-banking, land management (e.g., erosion control, sustainable forestry), HIV/AIDS defense and victim care (especially in Africa), and perhaps other entrepreneurial skills to help them market animal byproducts such as milk and calves, or feed.<sup>15</sup> The evidence from Nepal, in particular, suggests that the results of this additional training are often spectacular increases in long-term benefits, including major payoffs outside the primary recipient sites. But getting those benefits clearly requires serious evaluation of the training effort, and providing it has paid off handsomely.

We could run this situation through the TEC, to see whether we should do, and evaluate, the extra training (i.e., we could role-play the planner), or how we should phrase an RFP

---

<sup>14</sup> Send to [mjscriv1@gmail.com](mailto:mjscriv1@gmail.com) with the title line beginning TEC so that I won’t miss it.

<sup>15</sup> The original donated livestock was cattle and this is still the modal gift species, although goats, sheep, chickens, fish, bees, and even elephants have joined a dozen species in the Heifer Ark when they seem best suited to local circumstances. Hence the animal byproducts vary, and include milk, honey, eggs, and manure.

(request for proposal) to get it done (role-playing the manager tasked with commissioning the training), or how we should monitor the training (role-playing the contractor or the manager supervising the contractor), or decide whether the training has been successful—and if not, why not (role-playing an internal or external evaluator).

[This example became so long, in order to be useful to Heifer International country offices, that it became too long to fit into this report. It is planned to have it available to those interested, and sent out to the Years 4 through 7 country offices (and to other people who request it), in late 2011. Meanwhile, the best ‘short form of the TEC’ is not obtained by skipping some points, but simply by considering all of them carefully, given what is known about the program in the particular context, and *selecting an affordable subset that look as if they need the most attention.*]

---

**ENDNOTE: on the distinction between training and teaching.** A possible definition that makes the somewhat fuzzy distinction between them is: "Teaching typically aims to increase *cognitive content* (both knowledge and understanding); training typically aims to increase *the repertoire of specific cognitive or behavioral skills and competencies*—but there is some overlap." The main practical implication of this definition is, it seems to me, that it makes clear why *extensive complex practice* is so central to training and to the evaluation of training (*mentored* practice in training, and *monitored* practice in the evaluation of training). “Extensive’ here means that the practice goes well beyond the examples used to teach the material. The point may be better made by adapting a remark made about magic and cooking from a Calvin Trillin essay in the *New Yorker* recently: he said that you can learn exactly how the tricks of the great magicians and chefs are done from books and teachers, but you need a great deal of training before you can do them yourself. Logically speaking, this is a development of Gilbert Ryle’s famous distinction between ‘knowing that’ and ‘knowing how’—it is the distinction between knowing *how it’s done* (which takes education and is ‘knowing that’) and knowing *how to do it* (which takes training, i.e., ‘know-how’, although sometimes experience is enough, especially if monitored). Another way to put it is to say that education gives you an external kind of knowledge about a skill, and training

provides the internal view.

This distinction has some wider importance, too, since it comes close to the heart of two other distinctions of great importance (e.g., in curriculum theory or design and curriculum evaluation): (i) that between content wanted by *the academic* and that preferred by *the mechanic*, and (ii) that between *science* and *technology*. In the field of evaluation theory itself, theorists who favor the (program) theory-driven approach are, so to speak, the academics (addressing the questions Why and How it works), while those who think that's too far beyond the eponymous business of an evaluator are the mechanics, who favor addressing just Whether it works, and perhaps, How to make it work better. It's important to understand that the two overlap, sometimes considerably. Wittgenstein once said: if you want to become a good philosopher, spend some time as a car mechanic, as he did. (I found similar benefits from working as a word processor trouble-shooter in the old days when word processing required a dedicated machine.<sup>16</sup>)

Incidentally, you don't need a valid program theory to tell you how to improve a program—you only need evidence that supports the claim that the change will produce improvement, which may be a single sentence expressing a single causal claim, e.g., "Switching to File-Maker Pro from your present database software will greatly increase the payoff from staff use of your data."

---

<sup>16</sup> See *Word Magic: Evaluating and selecting word processing*, Scriven, M. (1987), Wadsworth and van Nostrand. Translated into Russian and also published by the Soviet State Publishing House (Moscow), with acknowledgment but without permission.