LEARNING + DEVELOPMENT
By Georgia L. Harris, NIST Weights and Measures Division

Introduction
As metrologists, we think we are good (even expert) at measurements. But, our subject matter expertise is most often in the world of science, and not necessarily in the world of education. There are a number of aspects of measuring training that we could consider in a study about measuring the effectiveness of training. What most often comes to mind in training evaluation is the course evaluation. However, instructors need to use assessment methods within the classroom environment to determine if and how well students are learning. Universities need to measure the effectiveness of instructors and course curricula. In the workplace, managers want to invest in courses that make their employees more productive and efficient. In laboratory accreditation according to ISO/IEC 17025, a laboratory must also have a way to measure the effectiveness of training efforts. As you can tell already, training evaluations need to include more than simple course evaluations. With all of these various aspects of measuring effectiveness, where do we start?

Reasons for Training
Why have training at all? Generally, employers who invest in the training want to see results. If the training can improve performance so that an organization can be more effective, sell more product, have a higher quality product, make better measurements, or achieve accreditation more easily, the investment can be justified. Trainers want to be able to demonstrate that they are helping students improve their knowledge, skills, and attitudes (because that is their business and they need to make the case for training!). Students want to be able to demonstrate their value to the organization by implementing new knowledge and skills on the job to contribute to the success of the organization and thus personal success. We all care!

Sometimes, the solution is not training. The real objective (think “root cause” for training) is usually performance improvement; that requires a three-way partnership among the organization, the trainer, and the student. Everyone plays a part in ensuring performance improvements. For example, if the employer doesn’t provide time or opportunities to implement what was learned in a training event, nothing will change and the training was a waste of everyone’s time and money. If the trainer doesn’t provide adequate instruction, learning activities, and opportunities to apply learning during the event, the student may not learn effectively. And if the student is going to class “because my manager said I needed to come to keep my job,” there isn’t a lot of motivation to learn and apply new concepts and skills later. For training to be worthwhile, barriers to application of new ideas or good measurement practices must be removed and learning must be reinforced on the job.

As an aside note: improving performance in specific ways is a key reason to have well-written learning objectives. All three parties want to know what the employee will be able to know or do differently after the training event.

In addition to having effective partnerships among all parties, sometimes the performance improvement solution is the removal of a barrier to try new ideas or the use of a simple checklist to follow to apply a complex procedure. Formal training needs analysis is another form of evaluation and assessment that looks at performance improvement options, but is beyond the scope of this article. It’s a topic that we should come back to in the future. For the sake of this article, let’s say that we’ve determined that a training event is the best solution and all three parties are fully engaged. Now, let’s talk more about assessing learning.

Measuring Organizational Performance and Effectiveness
Much has been written on measuring training effectiveness by Kirkpatrick, Phillips, and Clark. In fact, much of the training assessment literature builds on the works of Kirkpatrick. Both Phillips and Clark suggest modifications of the basic ideas presented by Donald Kirkpatrick in 1959. The bottom-line for these approaches is the ability to measure whether the training makes an impact and if employee performance improves over the longer term. Let’s look at their ideas.

“If you cannot measure it, you cannot improve it.”
Lord (William Thomson) Kelvin
(1824-1907)
Kirkpatrick’s Four Levels presents a series of evaluation levels as shown in Table 1. The table shows the levels and descriptions of what is to be measured. Table 1. Kirkpatrick’s Four Levels of Evaluation (0th and 5th Levels added by Phillips)

<table>
<thead>
<tr>
<th>Level</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Inputs</td>
</tr>
<tr>
<td>1</td>
<td>Reaction</td>
</tr>
<tr>
<td>2</td>
<td>Learning</td>
</tr>
<tr>
<td>3</td>
<td>Behavior</td>
</tr>
<tr>
<td>4</td>
<td>Results</td>
</tr>
<tr>
<td>5</td>
<td>Return on Investment</td>
</tr>
</tbody>
</table>

**Level One**
Level one seeks to answer: Did the participant “like the training”? Course evaluations are used at level one and are quick and relatively easy to collect and assess. They are also inexpensive and don’t take a lot of time or energy to evaluate. A good course evaluation will also uncover if there are problems or barriers to learning that need to be improved or corrected (such as poor audio quality, missing handouts). But, level one course evaluations are often called “smile sheets” because they assess whether students were happy when the course ended. Course evaluations are often written from the perspective of the training provider or instructor and may not address what concerns the student may want to address at the end of a training event. Reaction may also tell you how well the training experience (needs analysis, materials, instructor, and environment) met the student expectations.

**Level Two**
It is important to focus on whether learning objectives were presented and covered and whether the student feels like they were achieved during the course (in addition to the administrative questions a provider wants answered.) Level two seeks to answer: Did the participant “learn anything”? Assessments such as pre-tests and post-tests are often used to assess this; however a course evaluation can gather information from the participant about whether they were able to demonstrate that they learned new knowledge or skills during the event. Observations should also be used by the instructors to assess whether learning occurred and whether learning objectives were achieved. The instructional design process should consider how to assess whether learning occurs during the event. It is more challenging to capture and report on whether learning occurred in a training event. Key questions to ask might be:
- What knowledge was acquired?
- What skills were developed or enhanced?
- What attitudes were changed?
One way to measure this is to make sure the learning objectives, activities, and assessments are aligned and then make sure that the course evaluation asks specific questions about whether objectives were achieved. The demonstration of skills, compiling test results, and reporting on case studies are all ways that can be used to assess whether learning occurred. Another measure that can be captured at level two is whether the training experience was effective in enabling learning.

**Level Three**
Level three requires an effective partnership between the employee/trainee and the employer/supervisor because it answers the question: does the participant “use the knowledge/skills/ attitude on the job”? It requires observation and interviews over time to assess change, relevance of change, and sustainability of change in on-the-job performance. Sometimes, an employee already knows how to perform certain skills on the job and chooses not to do so. In this case, training is not likely the solution unless there is a component of attitude improvement designed into the learning event – so that the employee now wants to perform in a certain way on the job. This kind of measurement of behavior requires cooperation and skill of supervisors. Another output of level three is whether the training experience enables improved performance on the job. Level three helps assess whether the training was effective.

**Level Four**
Level four evaluations answer questions about: What impact does the training have on “the business result(s)?” One big challenge here is to demonstrate clear correlation between...
the training and the impact. Business measures need to be in place to assess before and after conditions and is often quite difficult. A key aspect to consider at level four is whether the training experience led to the impact or whether barriers were removed, incentives were provided, or even whether new employees implemented systems that were already in place and not being followed prior to the training. The dynamics of the working environment make it difficult to assess training at the fourth level.

Modifications and Criticisms
Since Kirkpatrick published his original model in 1959, other authors (e.g., Phillips) have suggested a possible fifth level, such as Return On Investment (ROI). Some authors believe that ROI can actually be incorporated into the Results level.

One of the criticisms of the Kirkpatrick model is that there is little direct correlation between Level 1 (Student Reaction) and Level 3 (Behavior – Application on the job). Students may leave a seminar and not enjoy a single part of it, thus providing negative feedback on the course evaluation. Yet, they may be fully successful in applying knowledge and skills on the job. Given that most course evaluations are only looking at Level 1 (recall: smile sheets), what can that tell us about how we need to measure training effectiveness?

Another criticism is that the model is upside-down! Clark suggests that we start with the end in mind and look at the required level of performance first. He suggests that we flip the model and work backwards by identifying:
1. the desired impact (outcome or result) that will improve the performance of the business;
2. the level of performance the learners must be able to do to create the impact;
3. the knowledge and skills they need to learn in order to perform; and
4. what they need to perceive in order to learn X (the need to learn).

Phillips also presents a version of this approach called the V model. This approach fits well with the reasons for training we stated earlier: performance improvement. But, just because we have a model that fits our reasons for training, it doesn’t make it any easier to measure the results! Two things are sure: we need to engage all parties in the learning process and Level 1 course evaluations don’t measure it. That doesn’t mean we should throw out Level 1 evaluations; they are still important to assess barriers and areas for corrective action and course improvements.

The International Association for Continuing Education and Training (IACET) references the Kirkpatrick levels of evaluation as a part of their ANSI/IACET standard. Here are a couple of interesting statistics I’ve noted. One is that Bassi and others discovered that 96% of companies surveyed used some form of the Kirkpatrick framework to evaluate training and development programs. Another is that McMurrer and others surveyed the American Society for Training and Development Benchmarking Forum to determine what percentage each of Kirkpatrick’s four levels is used in organizations:
- Level 1: 95% (target goal: 100%)
- Level 2: 37% (target goal: 70%)
- Level 3: 13% (target goal: 30%)
- Level 4: 3% (target goal: 20%)

Phillips indicates that we probably shouldn’t seek to assess each level at 100% due to the cost and benefit required, but suggests target goals as those given in parenthesis above. Given that we are looking for training effectiveness in the laboratory working to meet the requirements of ISO/IEC 17025 and that only 13% of organizations are using level three assessments, we probably don’t have many good examples of best practices. In fact, failure to apply new concepts, or evidence of applying concepts, is generally used as a measure of training (in)effectiveness. For example, “my technical audit found non-conformities; therefore my audit training was effective.” Given that we really want to measure impact, results, and performance improvement, and only 3% of organizations have implemented level four assessments, this suggests we need to do a better job of measuring!

Measuring Student Learning

Student Assessments
Assessing student learning is generally a level two assessment. The last train the trainer article in Metrologist presented ideas about aligning learning objectives, activities, and assessments. Instructor assessments of student learning are a critical part of good instructional design concepts and are one way to assess whether students have learned. Assessments are a critical part of the training event, and they assess student learning as demonstrated during the event. This level of assessment may be incorporated into course evaluations, but generally only as “self declaration”.

If we are looking for performance improvements, additional tools can aid the student and the employer in evaluating whether the learning was effective. For example, a checklist used by the instructor to assess whether a student has learned a topic can also be given to the student so they know how they will be assessed and self-assess. The same checklist can be shared with the employer so that they can review or observe the employee’s work back on the job to make sure all items are implemented on the job. Job aids such as forms and checklists may even be more effective than a course notebook or set of slides to take home from a training event.

Course Evaluations
Course evaluations most often determine whether students were happy with the training experience (hence the phrase “smile sheets”) at level one. They can also provide an assessment of whether the learning event provided a good environment, resources, and opportunities to begin applying concepts and skills. Given the concepts presented in the previous sec-
tion on the four levels of evaluation and return on investment for a training program, it is unlikely for course evaluations to effectively get to the heart of whether there will be performance improvement on the job. However, designing course evaluations to assess student learning is still a worthwhile goal. Students may self-declare their level of learning and intent to apply concepts back on the job, which is a start. Having questions such as “what concepts/skills will you apply back on the job?” reinforces the expectation that the reason for the training event is for the student/employee performance improvement on the job. It also helps the instructors and training program managers identify the key concepts, skills, or knowledge that students found were valuable and can be applied back on the job.

Feedback from the course evaluations can help with further training needs analysis, identify topics for measuring impact, and provide details about what improvement actions may be needed in the environment, resources, and instruction. Assessing the course evaluation feedback is an important measure for continual training improvement. Additional resources on course evaluations can be found in the NCSLI Training Library.

The Perfect Course Evaluation
I read an online a story recently about an instructor who always gets perfect evaluations! The gist of the story was that the instructor spends most of his time ensuring that his evaluations are perfect rather than whether students are learning anything. In fact, he really doesn’t teach anything, just prepares the students to answer the course evaluations at 100 percent! Perhaps you’ve attended a training event with this instructor? I hope not. Teaching-to-the-test efforts are probably not much better. Please remember that the purposes of the course evaluation are to assess the learning experience, to have a continual improvement process, and eventually to measure the impact of our efforts.

Conclusion
Measuring the impact and effectiveness of training from a system perspective requires a three-way partnership between the employer, the trainer/training organization, and the student/employee. Everyone has a vested interest in measuring effectiveness and improving training results and employee performance. There is a lot we can learn from professional trainers about measurement. Ensuring that our course, tutorial, and conference evaluations consider more than whether the students had an enjoyable time is essential.

2010–2011 Joe D. Simmons Scholarship Recipient
By: Feilong Lin, U. of North Carolina Charlotte

Currently I am a doctoral student and research assistant in the Center for Precision Metrology (CPM) in UNC Charlotte. I started to get into the field of precision manufacturing and measurement as early as my master’s study in 2004 in China when I focused on how the vibration affects the surface profile of an aspheric workpiece during high-speed grinding. Since then, I learned the importance of measurement in modern manufacturing, that is, you need be able to measure something before you know what exactly it is and can make it better. When I came to the States in the fall of 2007, I started to work with Dr. Stuart Smith and other colleagues in the CPM, which is a state of art research center for precision metrology. Through my study and research here, I become more aware of the importance of precision metrology in many aspects of society, e.g. basic manufacturing, the 25 nm chip fabrication process in the semiconductor industry, cell imaging, nano-material science, etc. Because my interests and majors in China are related to mechanical, automation, electrical and programming, I continue my studies in precision engineering under the metrology instrumentation discipline. I have developed skills in the fields of precision machine design, control, signal process, programming, and electrical circuit design during the implementation of projects including the scanning probe microscope project with a multi-disciplinary research team. This device, which is capable of nanometer positioning accuracy in high scan speed, is used by researchers in physics, mechanical and nano-science in UNC Charlotte. Upon graduation, my object is to find a job in the field of precision metrology where I can use the knowledge and skills I learned during my doctoral study to build better performance machines with respect to the resolution, bandwidth, accuracy etc. After several years build-up of my career, hopefully I can become an expert in precision metrology. The Simmons Scholarship is a great support for me to graduate and achieve my goal and I appreciate a lot this award.

feilong.lin@gmail.com