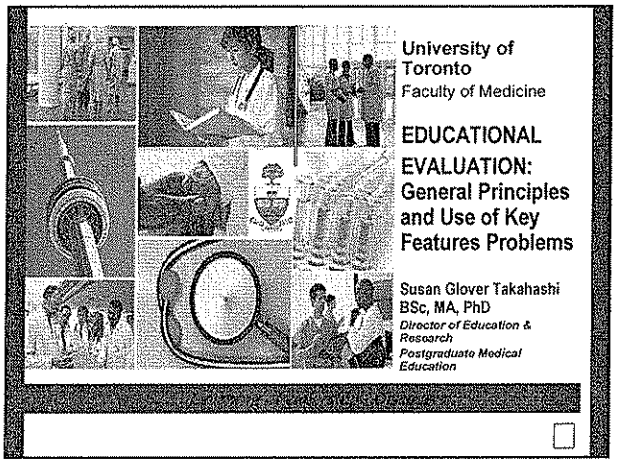


Overview of Presentation

Develop an understanding and practice in:

1. Principles of Assessment
2. LEARNER Assessment
3. PROGRAM Assessment
4. Key Feature Problems

1. Principles of Assessment

Definitions

- SEE HANDOUT
- Define your terms
- Shared understanding
- Point of view

Options & Issues

- Learner VS Program/system
- More is better VS.Good enough
- Formative VS Summative
- Higher stakes VS. Lower stakes
- Expert opinion VS Data
- Purpose of the educational object
- Mandate of organization

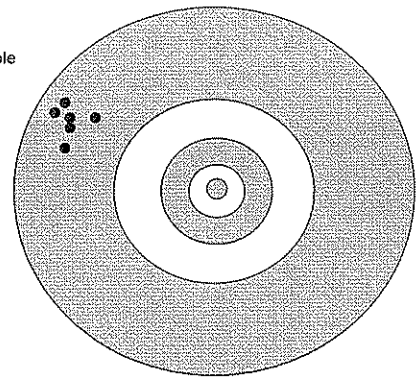
ASSESSMENT vs. *assessment*

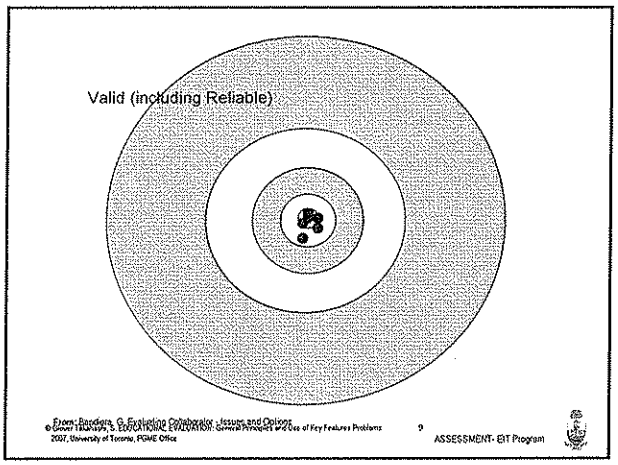
- Purpose of assessment----what are you measuring, why
- Principles of assessment
- Processes
- Standards

Principles of ALL assessments

1. Reliability of assessment = Reproducibility---><u>*accuracy of exam indicator*</u>
2. Validity➔➔ of inferences, information, & decisions (NOT of exam)---><u>*accuracy of decision indicator*</u>

Reliable





Validity

- Validity through: content
  criterion
  face
  construct
  reliability (administration, scaling, scoring, reporting)

Standard Setting

- Assumptions
  - Minimal standard?
  - Entry-level standard?
  - Advanced standard?
- Standard Setting for the sub parts
- Standard Setting for overall decision making

2. Learner Assessment

Assessing Learner?

- Learner assessment mechanisms must be:
  - consistent with educational purposes;
  - legally and professionally defensible;
  - equitably applied and enforced; and,
  - administratively and economically feasible
- What competencies (KSA are you assessing)

Elements of Competent Practice

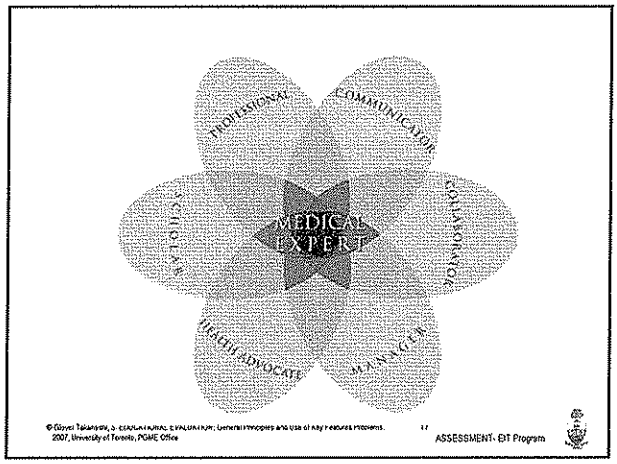
COMPETENT =
Capacity
+ Competencies
+ Context

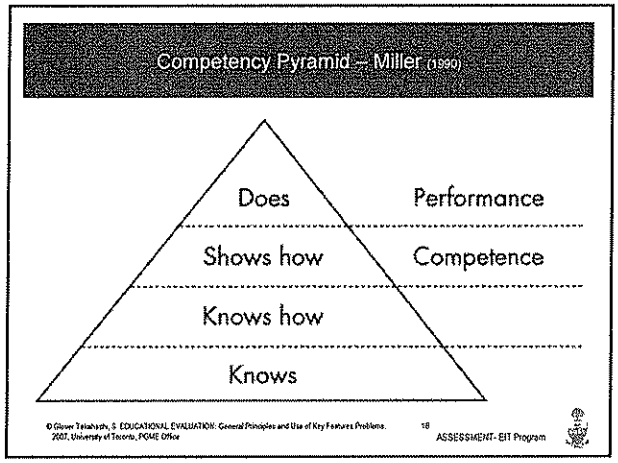
Competencies

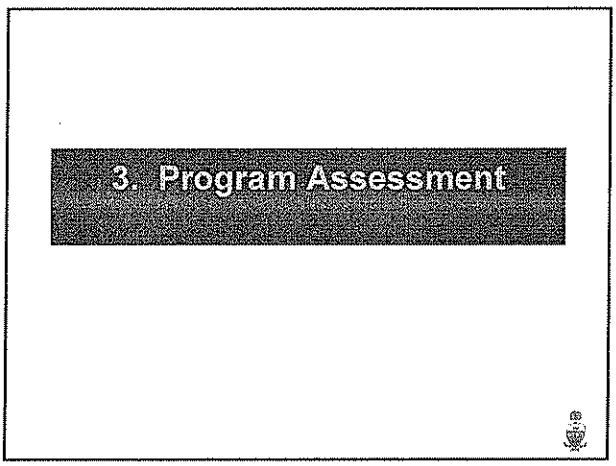
- Evaluate a small number of important competencies
- Describe observable behaviors
- Describe measurable behaviors
- Describe the outcome
- Ensure ease of understanding and use

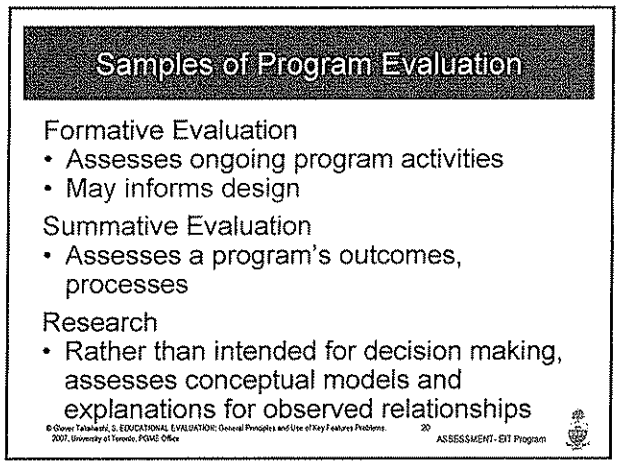
Learner Competency Outlines

- Educational accreditation agency
- Examination guideline
- Educational program
- Professional bodies
- Provincial-based standards?
- Expert opinion









Kirkpatrick Model

Level 1: Reaction

- Learners' reactions to the Program
  - The relevance of the objectives
  - What works well and what doesn't work well
    - The ability of the course to maintain interest.
    - The amount and appropriateness of interactive exercises
  - The *perceived* value and transferability to the workplace
    - Expectations and the extent to which they were met

Level 2: Learning

- What knowledge was acquired?
- What skills were developed or enhanced?
- What attitudes were changed?

Level 3: Performance

- E.g." Do people use their newly acquired learning on the job?"

Level 4: Impact

- "What impact has the training achieved?"
- These impacts can include such items as efficiency, moral, teamwork, etc.
  - **Quality training.** Measure a reduction in number of defects.
  - **Safety training.** Measure reduction in number or severity of accidents.

4. Key Features Problems

Key Features Problems

PURPOSE:
– Assessment of Ability to Problem Solve

ASSUMPTION
– Problem solving is situation specific
(i.e. Case specificity)

Measurement Issues for Key Features

- Content Validity
  - representative sample of problems
  - focus on the most important decisions
- Construct validity
  - assess knowledge application & what is really done

What kind of problems

- Real life situation
- What is the problem?
- What are the learning objectives
- What are the critical actions, unique challenges
- Focus the questions on the 'key features', discriminating elements to success, elements that make a difference

Testing tool options

- Multiple choice
- Fill in blanks
- Short answer
- True/false
- Matching

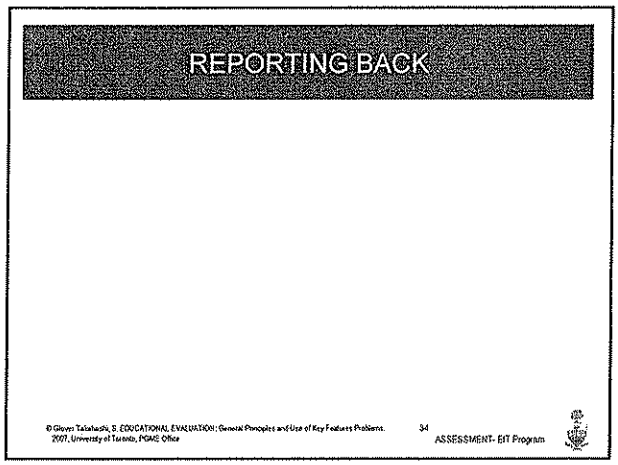
- OSCE
- Oral exam

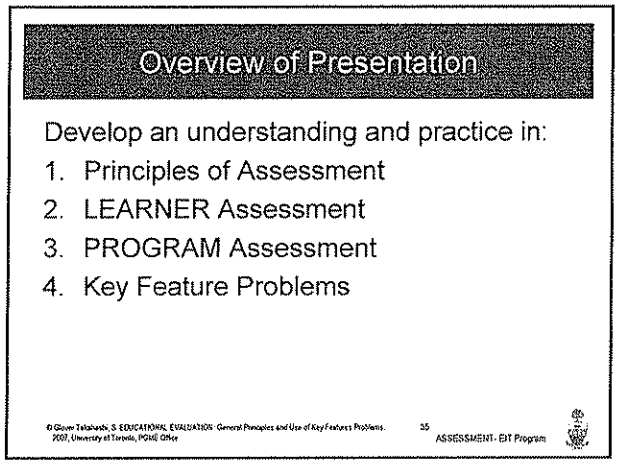
About Problems

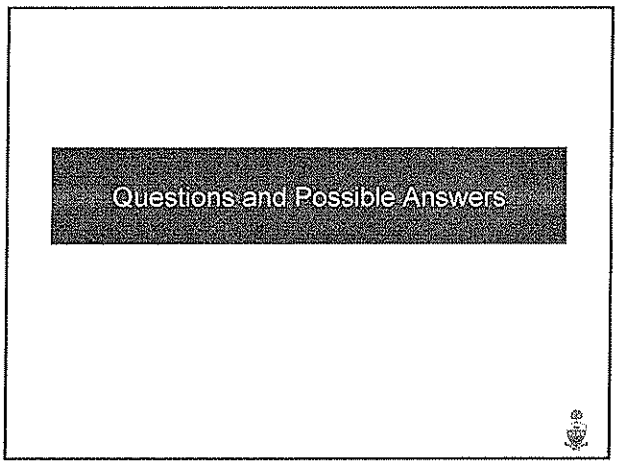
- Novices - Competent
  - Typical problems, typical presentations
  - Thorough, structured approach
- Competent - Experienced
  - Typical problems, typical + atypical presentations, rare problems
  - Focused or strategic approach
- Expert
  - Professional sophistication (ease, efficiency) in managing problems (i.e. typical problems, typical + atypical presentations, rare problems)
  - Selective approach to problem

PROBLEMS – pick one

1. Parallel parking
2. Launching a website
3. Creating a student registration database
4. Patient with a headache in emergency department
5. Patient's daughter unhappy father has been waiting to see doctor in clinic for 3 hours







SOURCE:

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

GLOSSARY of KEY TERMS

ACADEMIC CREDENTIAL ASSESSMENT, also called ***CREDENTIALLING,*** typically includes a paper review of documents such as degrees / diplomas, transcript of records and course materials that systematically evaluate their education to ensure that it is sufficiently similar to the established and published standard.

ACCREDITATION: The process of external quality review used in higher education to scrutinize colleges, universities, and higher education programs for quality assurance and *quality improvement*. Success results in an accredited institution and/or program. In some countries, it conveys institutional authority to offer specific programs. *(See info about some specific regions*)

- **South Africa:** Refers only to institutions and their authority to offer specific programs.
- **North America:** A collegial process of self-study and external peer review for quality assurance, accountability, and quality improvement of an academic institution or program designed to determine whether or not it has met or exceeded the published standards of its accrediting association and is achieving its mission and stated purpose.
- **Western Europe:** An evaluation and assessment of an institution or its programs in relation to its aims and objectives, its recognized standards, and its own goals. The assessors are looking primarily at the success of the institution in achieving its goals. Also refers to formal government authorization given to institutions to grant degrees. In Germany, normally refers to evaluation and assessment of the accreditation agencies that accredit only those programs leading to a B.A. or M.A.
- **United Kingdom:** the *QAA Code of Practice* on collaborative provision—a process by which an institution without its own degree-awarding powers is given wide authority by a university or other awarding institution to exercise powers and responsibility for academic provision. The awarding institution is ultimately responsible for the quality and standard of the award (qualification).

APPLICANTS: those who would like to have assessment of their substantial equivalency conducted.

ASSESSMENT: A diagnostic form of review of an individual's knowledge, skills or abilities. Sometimes the words assessment and evaluation are used interchangeably. Here the word assessment pertains to the individual and evaluation applies to a review of a program or organization.

COMPARABLE means that there is confidence that the qualifications are generally similar.

SOURCE:

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

COMPETENCIES are the knowledge, skills, and abilities obtained through formal, non-formal, or informal learning; ability to perform occupation-specific tasks and duties.

Often two 'levels' of competencies are described, key competencies and enabling competencies.

KEY COMPETENCIES are the important outcome objectives (i.e. ***what is to be achieved or performed***). Central to the accuracy of the competencies is the action verb.

Enabling Competencies are the sub objectives, or ***key ingredients to achieving the Key Competencies.***

COMPETENT refers to the skill level of a practitioner which meets or exceeds the minimal and ongoing performance expectations and demonstrates the requisite knowledge, skills, and abilities for safe and effective professional practice at the beginning of *and* throughout their career.

CONTEXT OF PRACTICE refers to the environment where practice occurs. To function safely and effectively professionals need to have an understanding of such local jurisdictional aspects as: jurisprudence, values and ethical framework, health system delivery systems, health policies.

COMPETENCY ASSESSMENT –is diagnostic form of evaluation of the individual's capacity to perform adequately in the:

- essential knowledge, skills or abilities,
- required roles, and
- necessary work context.

CONTEXT OF PRACTICE: environment where practice occurs. To function safely and effectively in the receiving jurisdiction's health practice context, professionals need to have an understanding of such local jurisdictional aspects as: jurisprudence, values and ethical framework, health system delivery systems, health policies.

CONTINUING COMPETENCE:
The ongoing ability of a practitioner to integrate and apply the knowledge, skills, judgements and interpersonal attributes required to practice safely and ethically in a designated role and setting.[1]

CREDENTIALLING, see Academic Credential Assessment.

CRITERIA: Standards for review that relate to the expectations about quality, effectiveness, financial viability, compliance with rules and regulations, outcomes, and sustainability.

---

[1] College and Association of Registered Nurses of Alberta. (2003) Continuing Competence Program. Available: http://www.nurses.ab.ca/contcomp/index.html

SOURCE:

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

JURISDICTION OF ORIGIN means the country, province or state jurisdiction in which a professional receives their profession specific education and other qualification which permit professional practice.

LICENSURE, see ***REGISTRATION.***

MOBILITY is the extent to which a worker is able to move from country or jurisdiction to another and to gain entry into a profession without undue obstacles or barriers.

PORTFOLIO ASSESSMENT is the systematic evaluation of formally presented documentation and other supporting evidence that demonstrates and provides validation of learning, and articulates the learning towards course or program requirements.

PROFESSION-FOCUSED CREDENTIALING looks at the nature and scope of the individual's documentary evidence to evaluate the similarity of the educational curriculum - i.e. scope of educational program, content, scope of key professional coursework domains, methods of instruction and evaluation, fieldwork or other requirements.

PROFICIENT refers to the skill level of a practitioner where the person performs a similar competency (i.e. to the competent practitioner) with enhanced ease and sophistication and a greater capacity to deal effectively with a wider range of complexity. The practitioner at the proficient level has additional capacity to perform HOW practice is demonstrated than the practitioner at the competent level.

PRIOR LEARNING ASSESSMENT (PLA), sometimes also called Prior Learning Assessment and Recognition (PLAR) involves the identification and measurement of skills and knowledge acquired outside formal educational institutions. Assessments are most often used to grant academic credit or determine eligibility to practice a trade or profession. Recognition is based on an assessment of skills and knowledge obtained through work and other life experience. Prior Learning Assessment and Recognition may also include determination of future goals and individual education needs.

QUALIFICATION RECOGNITION is a qualitative process of systematically evaluating the documentary evidence provided by a professional to determine if their qualifications are sufficiently similar to the established and published standard of the receiving country or jurisdiction. Qualification recognition can include academic credentialing assessment, prior learning assessment and portfolio assessment.

QUALIFICATIONS are the possession of knowledge, skills, and experience for entry to an educational program or practice in an occupation.

SOURCE:

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

QUALITY ASSURANCE: Planned and systematic review process of an organization or institution or program to determine that acceptable standards (e.g. scholarship and infrastructure) are being maintained and enhanced.

QUALITY IMPROVEMENT: The expectation that an organization or institution will have in place a plan to monitor and improve the quality of its programs.

RECEIVING JURISDICTION means the country, province or state jurisdiction to which a professional is arriving/immigrating and is seeking recognition of their qualifications.

RECIPROCITY: A system in which accrediting bodies acknowledge each other's accreditation or certification decisions, even though the criteria may not be identical. Recognition is based on the assumption of equivalence of standards and quality.

RECOGNITION: Acknowledgment that an individual, organization or institution meets or surpasses standards set by the entity conducting the recognition review.

RECOGNIZED BODIES: are those universities and other institutions with power to award diplomas, degrees or other educational credentials.

REGISTRATION, *also called* ***LICENSURE,*** refers to approval of an individual to practice a profession once minimal requirements are met. Approval generally granted through governmental organization or organization granted authority by government.

STANDARDS: The level of requirements and conditions that must be met by individuals, organizations, institutions or programs to be recognized (e.g. registered, accredited or certified) by a quality assurance or accrediting agency. These conditions involve expectations about quality, attainment, effectiveness, financial viability, outcomes, and sustainability. (See also criteria, quality)

SUBSTANTIAL EQUIVALENCY: Substantially equivalency is a term that infers that following a review, a finding that the individuals, organizations, institutions or programs under review is in all essential respects the equivalent of or similar to the comparator and meets accepted quality standards. Substantial equivalency of individuals infers that the review has established a reasonable confidence that individuals possess the requisite competencies.

SUFFICIENTLY SIMILAR is the degree to which the qualifications need to be the same or similar which has been established by the receiving country or jurisdiction. Sometime the receiving country or jurisdiction has an established requirement that the education is *substantially equivalent*.* Other times the requirement is that the qualifications are *comparable*.*

SOURCE:

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

REFERENCES

Assessment Strategies, (1998). Report on the development of the competency profile for the entry-level physiotherapist in Canada. Prepared for: Canadian Alliance of Physiotherapy Regulators, Canadian Physiotherapy Association, Canadian Universities Physical Therapy Academic Council: Toronto.

Bereiter, C, Scardamalia, M (1993). Surpassing Ourselves: An Inquiry into the Nature and Implications of Expertise. Chicago: Open Court Publishing.

Curry, L, Wergin, F., Eds. (1993).Educating professionals: Responding to new expectations for competence and accountability. San Francisco: Jossey-Bass.

Dreyfus, R, Dreyfus, S (1986). Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. New York:The Free Press.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. Journal of the American Medical Association, 287, 226-235.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. American Psychologist, 49, 725-747.

Eva, K (2002). The aging physician: Changes in cognitive processing and their impact on medical practice. Academic Medicine 77(10 suppl).S1-S6.

Glaser, R, Chi, M (1988). Overview. In Chi, M, Glaser, R, Farr, MJ (Eds). The Nature of Expertise. New Jersey: Lawrence Erlbaum Associates.

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

Glover Takahashi (2004). Stepping out of the Shadows: The Learning of Ethical Conduct through the "I" and "Eye" of Physiotherapists. Doctoral Thesis. University of Toronto.

Handfield, RS., Mann, KV., Challis, ME., Hobma, SO., Klass, DJ., McManus, IC., Paget, NS., Parboosingh, IJ., Wade, WB., & Wilkinson, TJ. (2002). Linking assessment to learning: a new route to quality assurance in medical practice. Medical Education, 36: 949-958, p. 949.

Hays, R.B., Jolly, B.C., Caldon, L.J., McCrorie, P., McAvoy, P.A., McManus, I.C., & Rethans, J-J. (2002). Is insight important? Measuring capacity to change performance. Medical Education, 36: 965-971.

SOURCE:

Glover Takahashi, S.(2007), A Substantial Equivalency Assessment Framework, Federation of Regulated Health Professions of Alberta.

Harris, I. B. (1993). New expectations for professional competence. In L. Curry & J. F. Wergin (Eds.), Educating professionals: Responding to new expectations for competence and accountability (pp. 17-52). San Francisco: Jossey-Bass.

Herold McIlroy, J., & Glover Takahashi, S. (2004). Expertise in physiotherapists: A model for practice in Canada. Unpublished manuscript.

Higgs, J, Bithell, C (2001). Professional expertise. In Higgs, J, Titchen, A (Eds). Practice Knowledge and Expertise in the Health Professions. Oxford: Butterworth-Heinemann.

Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. Academic Medicine, 76: S87-S89

Parry, S. (1996, July). The quest for competencies. Training, 33, Minneapolis. p. 48.

Patel, V, Arocha, JF, Kaufman, DR (1999). Expertise and tacit knowledge in medicine. In Sternberg, RJ, Horvath, JA (Eds) Tacit Knowledge in Professional Practice: Researcher and Practitioner Perspectives. New Jersey: Lawrence Erlbaum Associates.

Rodolfa, E., Bent, R., Eisman, E., Nelson, P., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. Professional Psychology: Research and Practice: 36, 347-354.

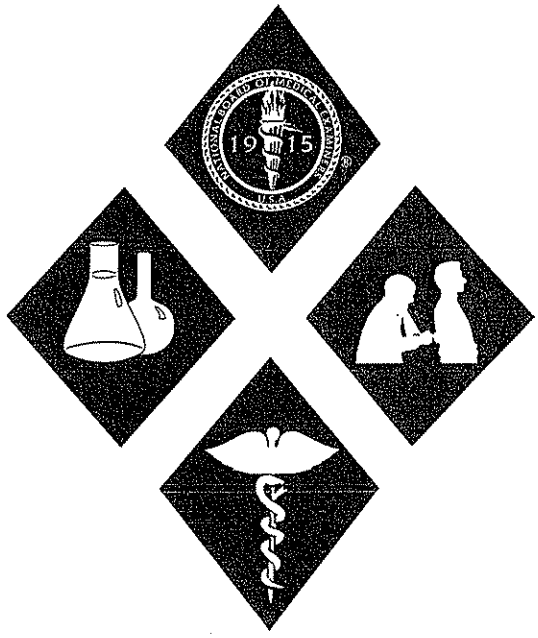
Schmidt, HG, Norman, G, Boshuizen (1990). A cognitive perspective on medical expertise: Theory and implications. Academic Medicine 65(10):611-621.

Schmidt, HG, Boshuizen, HPA (1993). On acquiring expertise in medicine. Educational Psychology Review 5(3):205-221.

Wass, V, van der Vleuten, C, Shatzer, J, Jones, R (2001). Assessment of clinical competence. The Lancet 357(March 24):945-949.

Constructing Written Test Questions For the Basic and Clinical Sciences

Third Edition
(Revised)

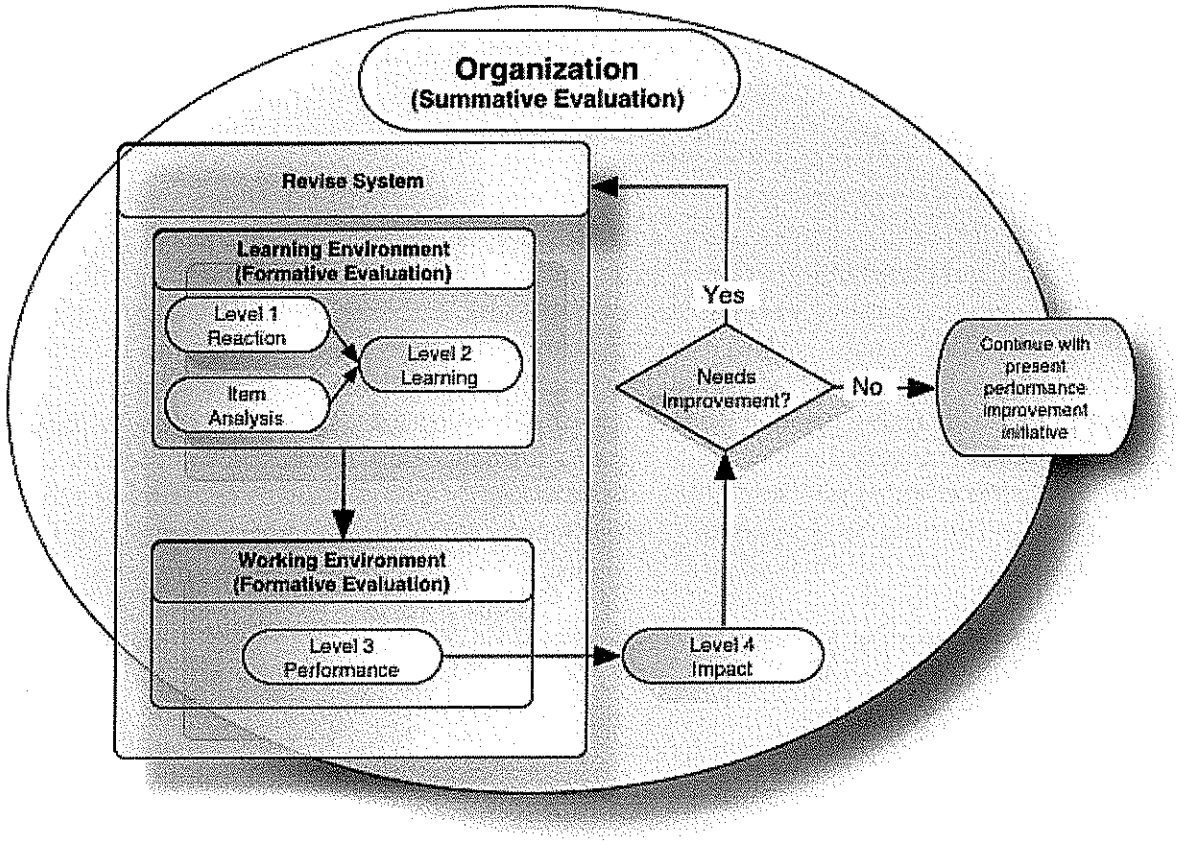


National Board of Medical Examiners
3750 Market Street
Philadelphia, PA 19104

http://www.nbme.org/publications/item-writing-manual.html

The Four Levels of Training Evaluation

Perhaps the best known training methodology is *Kirkpatrick's Four Level Evaluation Model* (1994) of reaction, learning, performance, and impact. The chart below shows how the evaluation process fits together:



Level One - Reaction

As the word implies, evaluation at this level measures how the learners react to the training. This level is often measured with attitude questionnaires that are passed out after most training classes. This level measures one thing: the learner's perception (reaction) of the course.

Learners are keenly aware of what they need to know to accomplish a task. If the training program fails to satisfy their needs, a determination should be made as to whether it's the fault of the program design or delivery.

This level is not indicative of the training's performance potential as it does not measure what new skills the learners have acquired or what they have learned that will transfer back to the working environment. This has caused some evaluators to down play its value. However, the interest, attention and motivation of the participants are critical to the

success of any training program. People learn better when they react positively to the learning environment.

When a learning package is first presented, rather it be e-learning, classroom training, CBT, etc., the learner has to make a decision as to whether he or she will pay attention to it. If the goal or task is judged as important and doable, then the learner is normally motivated to engage in it (Markus & Ruvulo, 1990). However, if the task is presented as low-relevance or there is a low probability of success, then a negative effect is generated and motivation for task engagement is low.

This differs somewhat from Kirkpatrick. He writes, *"Reaction may best be considered as how well the trainees liked a particular training program"* (1996). However, the less relevance the learning package is to a learner, then the more effort that has to be put into the design and presentation of the learning package. That is, if it is not relevant to the learner, then the learning package has to "hook" the learner through slick design, humor, games, etc. This is not to say that design, humor, or games are not important. However, their use in a learning package should be to promote the "learning process," not to promote the "learning package" itself. And if a learning package is built of sound design, then it should be help the learners to fix a performance gap. Hence, they should be motivated to learn! If not, something went dreadfully wrong during the planning and building processes! So if you find yourself having to hook the learners through slick design, then you probably need to reevaluate the purpose of the learning program.

For more information on reaction, see Self-System.

Level Two - Learning

This is the extent to which participants change attitudes, improve knowledge, and increase skill as a result of attending the program. It addresses the question: *Did the participants learn anything?* The learning evaluation require post-testing to ascertain what skills were learned during the training. In addition, the post-testing is only valid when combined with pre-testing, so that you can differentiate between what they already knew prior to training and what they actually learned during the training program.

Measuring the learning that takes place in a training program is important in order to validate the learning objectives. Evaluating the learning that has taken place typically focuses on such questions as:

- What knowledge was acquired?
- What skills were developed or enhanced?
- What attitudes were changed?

Learner assessments are created to allow a judgment to be made about the learner's capability for performance. There are two parts to this process: the gathering of information or evidence (testing the learner) and the judging of the information (what does the data represent?). This assessment should not be confused with *evaluation.*

Assessment is about the progress and achievements of the individual learners, while evaluation is about the learning program as a whole (Tovey, 1997, p. 88).

Evaluation in this process comes through the learner assessment that was built in the design phase. Note that the assessment instrument normally has more benefits to the designer than to the learner. Why? For the designer, the building of the assessment helps to define what the learning must produce. For the learner, assessments are statistical instruments that normally poorly correlate with the realities of performance on the job and they rate learners low on the "assumed" correlatives of the job requirements (Gilbert, 1998). Thus, the next level is the preferred method of assuring that the learning transfers to the job, but sadly, it is quite rarely performed.

Level Three - Performance (behavior)

In Kirkpatrick's original four-levels of evaluation, he names this level "behavior." However, behavior is the action that is performed, while the final results of the behavior is the performance. Gilbert said that performance has two aspects -- behavior being the means and its consequence being the end (1998). If we were only worried about the behavioral aspect, then this could be done in the training environment. However, the consequence of the behavior (performance) is what we are really after -- can the learner now perform in the working environment?

This evaluation involves testing the students capabilities to perform learned skills while on the job, rather than in the classroom. Level three evaluations can be performed formally (testing) or informally (observation). It determines if the correct performance is now occurring by answering the question, "Do people use their newly acquired learnings on the job?"

It is important to measure performance because the primary purpose of training is to improve results by having the students learn new skills and knowledge and then actually applying them to the job. Learning new skills and knowledge is no good to an organization unless the participants actually use them in their work activities. Since level three measurements must take place after the learners have returned to their jobs, the actual Level three measurements will typically involve someone closely involved with the learner, such as a supervisor.

Although it takes a greater effort to collect this data than it does to collect data during training, its value is important to the training department and organization as the data provides insight into the transfer of learning from the classroom to the work environment and the barriers encountered when attempting to implement the new techniques learned in the program.

Level Four - Results

This is the final results that occur. It measures the training program's effectiveness, that is, "What impact has the training achieved?" These impacts can include such items as monetary, efficiency, moral, teamwork, etc.

While it is often difficult to isolate the results of a training program, it is usually possible to link training contributions to organizational improvements. Collecting, organizing and analyzing level four information can be difficult, time-consuming and more costly than the other three levels, but the results are often quite worthwhile when viewed in the full context of its value to the organization.

As we move from level one to level four, the evaluation process becomes more difficult and time-consuming, however,it provides information that is of increasingly significant value. Perhaps the most frequently type of measurement is Level one because it is the easiest to measure. However, it provides the least valuable data. Measuring results that affect the organization is considerably more difficult, thus it is conducted less frequently, yet it yields the most valuable information.

Each evaluation level should be used to provide a cross set of data for measuring training program.

The first three-levels of Kirkpatrick's evaluation -- Reaction, Learning, and Performance are largely "soft" measurements, however decision-makers who approve such training programs, prefer results (returns or impacts). That does not mean the first three are useless, indeed, their use is in tracking problems within the learning package:

- Reaction informs you how relevant the training is to the work the learners perform (it measures how well the training requirement analysis processes worked).
- Learning informs you to the degree of relevance that the training package worked to transfer KSAs from the training material to the learners (it measures how well the design and development processes worked).
- The performance level informs you of the degree that the learning can actually be applied to the learner's job (it measures how well the performance analysis process worked).
- Impact informs you of the "return" the organization receives from the training. Decision-makers prefer this harder "result," although not necessarily in dollars and cents. For example, a recent study of financial and information technology executives found that they consider both hard and soft "returns" when it comes to customer-centris technologies, but give more weight to non-financial metrics (soft), such as customer satisfaction and loyalty (Hayes, 2003).

Note the difference in "information" and "returns." That is, the first three-levels give you "information" for improving the learning package. While the fourth-level gives you "impacts." A hard result is generally given in dollars and cents, while soft results are

more informational in nature, but instead of evaluating how well the training worked, it evaluates the impact that training has upon the organization. There are exceptions. For example, if the organizational vision is to provide learning opportunities (perhaps to increase retention), then a level-two or level-three evaluation could be used to provide a soft return.

This final measurement of the training program might be met with a more "balanced" approach or a "balanced scorecard" (Kaplan & Norton, 2001), which looks at the impact or return from four perspectives:

- Financial: A measurement, such as an ROI, that shows a monetary return, or the impact itself, such as how the output is affected. Financial can be either soft or hard results.
- Customer: Improving an area in which the organization differentiates itself from competitors to attract, retain, and deepen relationships with its targeted customers.
- Internal: Achieve excellence by improving such processes as supply-chain management, production process, or support process.
- Innovation and Learning: Ensuring the learning package supports a climate for organizational change, innovation, and the growth of individuals.