# Embedded Evaluation:
## Blending Training and Assessment

by
**Richard J. McCowan**
**Sheila C. McCowan**
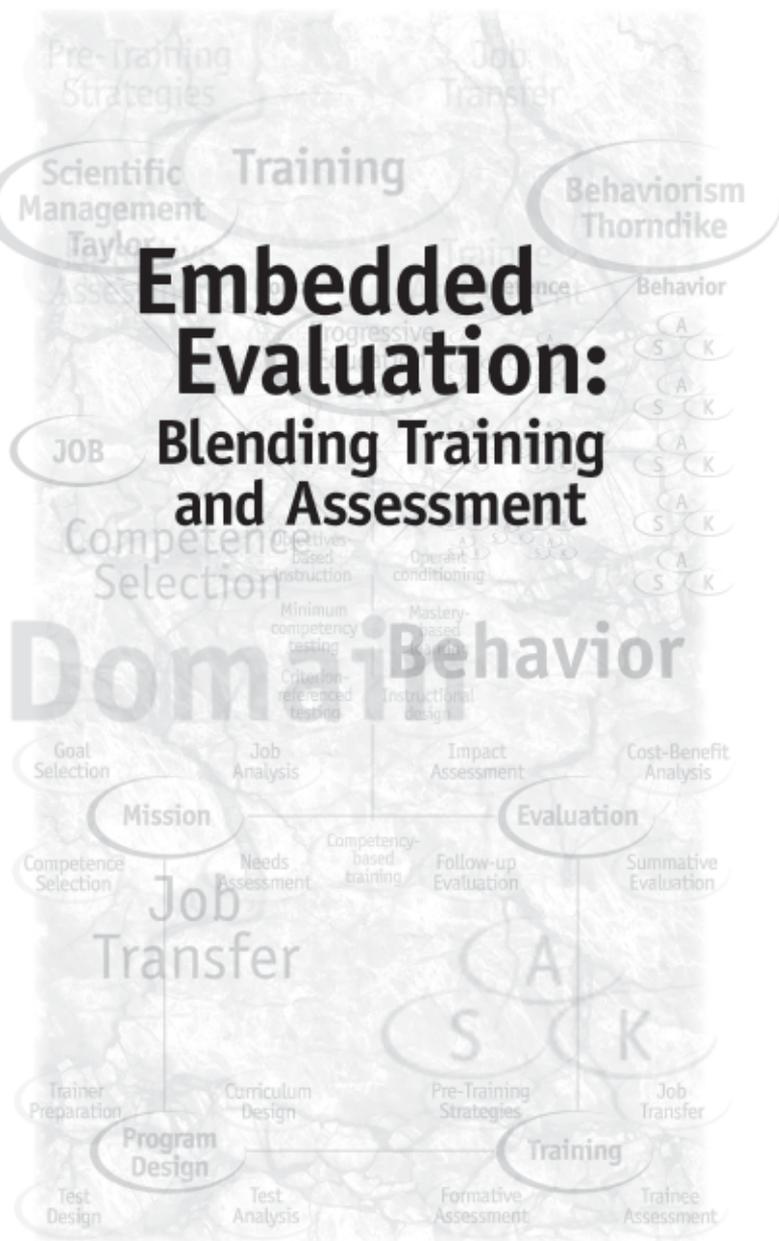
# Embedded Evaluation:
## Blending Training and Assessment

## CDHS

*Center for Development
of Human Services*

# Foreword

The embedded evaluation model described in this monograph was developed for the New York State Child Welfare/Child Protection Services Common Core training program. The Center for Development of Human Services developed this new, skill-based curriculum under the direction of the Bureau of Training and Workforce Development of New York State Office of Children and Family Services. Thomas Morton of the Child Welfare Institute in Atlanta, Georgia designed the curriculum model and developed the competencies on which the program is based.

This evaluation design provides a mechanism whereby skills can be assessed as part of the training program as situated assessment activities incorporated into the curriculum. This monograph describes the learning theories on which embedded evaluation is based and illustrates how this evaluation design can be implemented within a training program.

# Introduction

This paper describes an evaluation model called embedded training evaluation (ETE) which is used to evaluate different types of training ranging from basic skills through constructed learning. ETE is particularly effective in evaluating high-level performance skills using items that require examinees to construct responses. It can assess inservice and on-the-job training, either independently or in combination with regular classroom instruction. It uses using constructed tasks integrated with the curriculum and blends instruction with evaluation.

ETE utilizes curriculum-related assessment tasks situated in realistic, job-related settings within the classroom. It includes evaluation tasks that are a natural part of training, rather than tests that are external, the training program. Evaluation tasks are a natural part of training rather than an external or foreign element. Embedded tasks take less time from the instructional process and place fewer demands on trainers and trainees (Wiley & Haertel, 1996). ETE strategies are authentic because trainees demonstrate what they can do as workers in a job-like setting. When properly designed, assessments are indistinguishable from instruction.

ETE takes place in a setting analogous to the workplace using a training design similar to an assessment center. It incorporates key aspects of a cognitive apprenticeship that allows trainees to emulate processes experts use to solve job-related problems. Rather than using conventional testing and assessment practices, ETE uses skill-based items derived from the curriculum that combine assessment and training into a single, coherent package. When coordinated with on-the-job experiences, ETE is a powerful training tool. It provides a context appropriate for assessing constructive knowledge, but does not preclude testing objective and cognitive knowledge. Since this analytic, controlled process makes evaluation an integral part of training, it is a cost-effective, controlled alternative for assessing trainee competence.

## Background

A brief review of three major learning theories will place ETE into perspective as a design that facilitates assessment of constructed tests, but not preclude the use of items that assess knowledge at the behavioral and cognitive levels. The following summary discusses behavioral, cognitive, and constructive learning theories.

Behaviorism, cognitive science, and constructivism are alternative conceptions of learning ranging from externally mediated reality to internally mediated reality (Jonassen, 1991b). No one is completely objective or subjective during learning. When observing an event and communicating to others, an observer starts from a personal point of view.

Subjectivism is reduced when the observer communicates with others who do not have the same frame of reference. This shared reality is a compromise between objectivism and subjectivism.

Behaviorists regard learning as passive, primarily automatic responses to external factors in the environment.

Cognitivists think that learning requires abstract, symbolic representations by individuals.

Constructivists view knowledge as relativistic according to time and setting and constructed by each learner

## Behaviorism

At the turn of the century, Edward L. Thorndike dominated learning theory. In *Animal Intelligence (*1898) he described his basic concepts and applied his theory of connectionism to human beings in *Educational Psychology* (1903). He and other behaviorists, such as Clark Hull, Neil Miller, and Albert Bandura, left an indelible mark on education and training.

Thorndike said that learning was based on trial and error associations strengthened or weakened through use or disuse and by activity and experience. Learners were passive entities who merely responded to environmental stimuli. The human mind was a machine with hundreds of thousands of individual connections, each con-

taining a message with no necessary relationship with other messages (Kliebard, 1986). Unobservable internal states of mind had no effect on learning.

Thorndike's classic example of connectionism was a cat that learned to escape from a "puzzle box." After many trial and error efforts, the cat learned that pressing a lever opened a door allowing it to escape from the box. A connection was established because the pairing occurred was rewarded (law of effect) and formed a single sequence (law of readiness).

Thorndike believed in quantifying behavior:

> Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity, as well as its quality. Education is concerned with changes in human beings; a change is a difference between two conditions, each of these conditions is known to us only by a the products produced by it — things made, words spoken, acts performed, and the like. To measure any of these products means to define its amount in some way so that competent persons will know how large it is, better than they would without measurement. (Thorndike, 1918, p. 16)

Behaviorists assumed that "the world is completely and correctly structured in terms of entities, properties, and relations" (Duffy & Jonassen, 1991, p. 8). Knowledge is stable and independent of the individual because essential properties of objects are known and relatively unchanging. People use rational, systematic rules to draw logical conclusions. Behaviorists believe that learning is behavioral change due to stimulus-response reinforcements based on external reality (Jonassen, 1991b; Streibel, 1986).

B. F. Skinner developed the theory of operant conditioning. Learning was a sequence of stimulus/response behaviors by a learner, and instruction modified behavior by providing conditions that reinforced learners who exhibited correct responses. Learners linked responses to lower-level skills and created chains of higher-level

skills. Instructional design required the identification and sequencing of each step required to perform the desired behavior.

Skinner assumed that the mind is an inaccessible "black box" that can only be understood by observing overt behavior. He studied the feedback loop that connects overt behavior to stimuli that activate the senses. Skinner (1961) said that inner entities did not "cause" because the ultimate cause of behavior was external to the organism. He felt that behavior could be predicted, controlled, and modified without concern about the inner man.

## Cognitive Psychology

Cognitive psychologists were concerned with unobservable strategies used by people to solve problems. They focused on the cognitive structures (e.g., schemata and heuristics) that underlie phenomena such as problem-solving and transfer ability. They classified the mind as a reference tool that mediates between people and reality. People learn by using metacognitive strategies based on effective, sequential mental activities (Jonassen, 1991b).

Gagné (1965; 1968) was a cognitive psychologist who exerted a major influence on instruction. Based on his military training research during World War II, he developed an instructional design system that supplemented traditional learning principles by analyzing learning tasks into discriminations, classifications, and response sequences based on prerequisites for learning a more complex task. He maintained:

> The basic principles of design consist of: (a) identifying the component tasks of a final performance; (b) insuring that each of these component tasks is fully achieved; and (c) arranging the total learning situation in a sequence which will insure optimal mediational effects from one component to another (Gagné, 1962b, p. 88)

Gagné stressed that instructional design was concerned with

> such things as *task analysis, intratask transfer, component task achievement,* and *sequencing* . . . These principles are not set in opposition to the traditional principles of learning, such as reinforcement, differentiation of task elements, familiarity, and so on, and do not deny their relevance, only their *relative importance*. They are, however, in complete opposition to the previously mentioned assumptions [that] "the best way to learn a task is to practice the task." (Gagné, 1962b, p. 88, emphasis in original)

Gagné expanded on concepts originated by Thorndike and Skinner by adding task analysis of the desired performance, hierarchical sequencing of subordinate knowledge and skills, and assessment of training performance on related training outcomes (Gagné, 1962a; 1965; 1968). He noted that "Analysis of a topic begins with the statement of the terminal objective—the performance or performances one expects the student to be able to exhibit after the learning topic has been completed" (Gagné, 1965, p. 245). Gagné significantly influenced curriculum design by sequencing instructional tasks based on their relative complexity with simpler components treated as prerequisites for more complex tasks. Although his framework focuses on intellectual skills, the theory has been used to design instruction in all domains (Gagné & Briggs, 1974).

## Constructivism

Two major figures – Jean Piaget and Lev Vygotsky – laid the foundation for contemporary constructivism. Piaget's developmental constructivism differed from Vygotsky who stressed the significance of social interaction as the primary mediator in learning. As noted later in this paper, ETE draws directly from Vygotsky's theory.

### Piaget and Developmental Constructivism

Piaget observed children to discover how they learned to know their world. He asked standard questions to follow a child's train of

thought because he believed that the spontaneous comments of children would reveal their logic and reasoning. Piaget concluded that intellectual development was influenced by the interaction of heredity and environment. As children developed and interacted with the world, they invented and reinvented knowledge. He thought that cognitive growth was an extension of biological growth governed by the same laws and principles and argued that intellectual development controlled all aspects of emotional, social, and moral development (London, 1988).

Piaget provided the foundation for modern day constructivism. He maintained that intelligence required organization and adaption. People organize thoughts so they make sense, separating more important from less important thoughts and connecting ideas. At the same time, they adapt new ideas from new experiences through assimilation and accommodation

### Piaget and Stages of Intellectual Development

Piaget discovered that the way children think and reason changes at different periods in their lives. He maintained that everyone passes through four qualitatively distinct, invariant stages that cannot be skipped or reordered. Normal children pass through these stages in the same order with some variability in the ages at which they attain each stage.

| Stage | Age Attained | Cognitive Task Mastered |
|---|---|---|
| Sensorimotor | birth to 2 years | concrete objects |
| Preoperational | 2 to 7 years | symbols |
| Concrete operational | 7 to 11 years | classes, relations, numbers and reasoning |
| Formal operational | 11 years and up | abstract thinking |

Piaget asserted that learning requires the construction and reconstruction of knowledge. A child must act on objects to develop knowledge because the mind organizes reality and acts upon it (Sigel, 1977). Children cannot learn until they develop specific prerequisites through maturation (Brainerd, 1978). The ability to learn cognitive concepts requires attainment of the appropriate stage of intellectual development. Children who function below that stage

cannot learn concepts of a higher stage. The dual process of assimilation-accommodation enables a child to form schema. Assimilation adds new information, while accommodation changes and adjusts new ideas to the existing cognitive organization (Berger, 1978).

Intellectual growth involves three fundamental processes:

Assimilation incorporates new events into pre-existing cognitive structures.

Accommodation changes existing cognitive structures to accommodate new information.

Equilibration strikes a balance between a person and the environment (i.e., between assimilation and accommodation).

When a child experiences a new event, disequilibrium sets in until the new information is assimilated and accommodated into revised schema to attain equilibrium. The establishment of equilibrium varies with the developmental level and the type of problem. Equilibrium is the major factor in explaining why some children develop logical intelligence more quickly than others do (Lavatelli, 40).

**Vygotsky and Social Constructivism**

The theory of embedded evaluation corresponds closely to Vygotsky's sociocultural theory. Although Vygotsky focused on children, his insights apply to adult learning (Tharpe & Gallimore, 1988). As learners participate in collaborative activities, they acquire new knowledge and strategies. Vygotsky (1978) observed that pointing a finger begins as a meaningless, grasping motion. Over time, as people react to the gesture, it becomes a movement with meaning and an interpersonal connection. Vygotsky maintained that human activity on the social and individual planes is mediated by semiotics which are tool and signs that facilitate co-generation of knowledge and problem-solving (Vygotsky, 1981).

Vygotsky stressed that learning takes place in a zone of proximal development (ZPD) defined as the distance between levels of actual and potential development. Actual development refers to ac-

complishments someone demonstrates independently, while potential development refers to areas in which someone needs assistance. By building on experience and providing moderately challenging tasks, instructors provide intellectual "scaffolding" that helps people learn and develop (Roblyer, Edwards, & Havriluk, 1997, p. 70). The ZPD provides a setting for interactions between experts and learners. Experts model solutions, help learners find solutions to problems, and monitor progress (Tharpe & Gallimore, 1988). Vygotsky (1962) was convinced that cognitive development occurred when people solved problems jointly. When experts and novices interact, novices can participate in forms of interaction that are beyond their competence when acting alone. Learners operate within constraints provided by the experts, but use words and other artifacts in ways that exceed their current comprehension. Cazden (1981) emphasized this point when she wrote of "performance before competence" in referring to mechanisms of language and cognitive development.

Vygotsky said that "the social dimension of consciousness is primary in time and fact" (Vygotsky, 1978, p. 30). The social world exerted the primary influence over cognitive development through a structure of cultural heritage. The relationship between the individual and the social environment was dynamic, and all aspects of personal development, including higher mental functioning, originate in society. In his genetic law of development, Vygotsky noted:

> Every function in the cultural development of the child comes on the stage twice, in two respects: first in the social, later in the psychological, first in relations between people as an interpsychological category, afterwards within the child as an intrapsychological category . . . All higher psychological functions are internalized relationships of the social kind, and constitute the social structure of personality (Valsiner, 1993, p. 67).

Vygotsky differed from Piaget in noting that "maturation is viewed as a precondition of learning but never the result of it" (Vygotsky, 1978, p. 80). Instead, he proposed:

> Learning awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment and with his peers. . . learning is not development; however, properly organized learning results in mental development and sets in motion a variety of developmental processes that would be impossible apart from learning. Thus learning is a necessary and universal aspect of the process of developing culturally organized, specifically human, psychological functions (Vygotsky, 1978, p. 90).

## Contemporary Constructivism

Constructivism evolved from cognitive learning theory which, in turn, evolved from behaviorism. Behaviorism and cognitive learning theories emphasize objects, while constructivism focuses on how people construct knowledge based on prior experience (Winn, 1990; Duffy & Jonassen, 1991). Three principles of constructivism describe appropriate instructional behavior:

(1) Teach where the students are.

(2) Base instructional decisions on the changing needs of students.

(3) Build a supportive learning environment to capitalize on student ability to construct knowledge.

Constructivist theorists maintain that people actively construct personal ways of thinking based on innate capacities and experience (Molenda, 1991). Learners are active participants who build knowledge from individual experiences. Piaget (1954) contended that people construct meaning from experience by accommodating or assimilating experiences. When experiences do not correspond, disequilibrium occurs. This forces people to create new mental schema that allow them to understand these experiences. Constructivists

contend that people learn best by working on meaningful tasks in realistic contexts (Duffy & Jonassen, 1991). With complex problems, they learn most effectively in settings where they explore multiple viewpoints. People who have difficulty understanding new experiences should be given opportunities to experience disequilibrium. These opportunities clarify understanding more effectively than other alternatives such as persuading, bullying, cajoling, or describing flaws in understanding (Brooks, 1990).

Fabricius (1983) modified Piaget's schema theory in noting that "reality becomes the phenomena we experience through construction." Constructed knowledge is "a set of socially negotiated understandings of the events and phenomena that comprise the experienced universe" (Tobin & Tippins, 1993, p. 4).

Children use experience and socialization to construct "common sense" views of the world (Driver, Asoko, Leach, Mortimer, & Scott, 1994). Constructivism is a philosophical position that views reality as an independent mental construction that is the basis of a revised world view (Saunders, 1992). Steffe (1990) noted that "Constructivists view learning as the adaptations children make in their functioning schemes to neutralize perturbations that arise through interactions with our world."

Wheatly (1991) described two learning principles based on constructivist theory.

> (1) People do not receive knowledge passively, but actively construct it. Ideas cannot be transmitted directly into someone's mind. Individuals must construct these meanings.

> (2) Cognition allows people to organize experiences. They do not find truth, but construct reasonable explanations for experience.

Constructivism utilizes situated learning where "knowledge is created and made meaningful by the context in which it is acquired" (Farmer, Buckmaster, & LeGrand 1992, p. 46). Situated learning requires authentic activities guided by expert practitioners set in a

culture of practice (Billett, 1994). Comparisons of the performance of novices and experts demonstrate that experts organize knowledge and recognize patterns to solve problems in new situations (Glaser & Chi, 1988) Experts develop sophisticated cognitive structures that help them to solve new problems easily based on prior experience. Novices cannot solve problems as well as experts because they have limited knowledge. Effective instructional designs help novices develop the knowledge and mental structures required for expert performance (Gagné, Briggs & Wager, 1992; Glaser, 1990).

The most effective training occurs in situation learning that involves practice and reinforcement in work-related contexts (Billett, 1993). "The focus of instruction should be on the individual's active construction of knowledge" (Stevenson 1994, p. 29) in which the essential role of trainers is "to facilitate construction of knowledge through experiential, contextual, and social methods in real-world environments" (Lynch 1997, p. 27). Constructive instruction focuses on the learning process rather than the instructional process (Stevenson, 1994).

Constructivism emphasizes the ability to solve real-life, practical problems. Learners work in cooperative groups focusing on projects that require solutions to problems, rather than instruction that requires learning specific content skills. The role of is to organize resources and guide learners who set personal goals and teach themselves (Roblyer, Edwards, and Havriluk, 1997).

Constructivist learning includes reflective thinking and authentic activities, including learner collaboration, consideration of multiple perspectives, and access to content experts who can model domain-specific skills. (Grabe & Grabe, 1998). It maintains that people have an innate drive to make sense of the world. Instead of absorbing knowledge passively, learners actively construct knowledge by integrating new information and experiences. Schemata used by learners include procedures, techniques, knowledge, attitudes, and values that integrate new experience into a revised worldview.

Knowledge construction involves functional and social contexts. Research supports the concept that learning involves unique, subjective interpretations and interaction with others (Johnson and

Thomas 1994). Useful learning requires intrinsic motivation based on a desire to construct meaning. Constructivist trainers coach students to participate in active inquiry and discover underlying meanings. Conversely, competency-based training, which originates from cognitive psychology emphasizes the dissemination of selected, sequenced knowledge and the measurement of how well learners acquire this information. Constructed training activities, such as situated learning, simulations, cognitive apprenticeships, and embedded evaluation, help trainees transfer learning. As trainees acquire skills through practice, trainers decrease assistance and allow learners to internalize information and construct personal knowledge (Farmer, Buckmaster, and LeGrand 1992).

Research on how people learn in the workplace demonstrates the effectiveness of constructivist learning. Practitioners in several professions revealed that they learned how to deal with ill-defined, complex, or risky situations by having experts model how to deal with the situations and subsequently guide them as they practiced these skills (Farmer, Buckmaster, & LeGrand, 1992). Novice coal miners learned best when they interacted with expert miners and novice workers during work. The workplace culture shaped unique personal construction of understanding. Workers valued direct instruction only when it provided explicit information they were unlikely to learn from other workers (Billett, 1993).

The major instructional task is to organize experiences that allow learners to develop competence. Instructors model, mediate, diagnose, and scaffold instruction as they coach learners through higher skill levels. The learning environment includes: authentic activities and experiences in actual or simulated contexts that resembles the workplace. Performance assessment uses items that reflect a level of construction appropriate to measure the skills in which people were trained (Billet, 1993).

The literature describes three views of constructivism including radical, moderate, and ration. ETE is based on the rational model.

> (1) Radical constructivism is highly subjective. In the extreme, it rejects realism (Goodman, 1984; Molenda, 1991). Radical constructivists believe that "What we

*know* to be real is the result of historical and social processes of meaning-making, language-making, and symbol-system making. The social construction of reality applies to our knowledge of physical reality, as well as to our knowledge of social reality" (Streibel, 1986, p. 138). People have different viewpoints because they interpret information differently. "There are many ways to structure the world and many meanings or perspectives for any event or concept. There is not a correct meaning that we are striving for" (Duffy & Jonassen, 1991, p. 8). Experience dictates understanding, schools should provide experiences that help students understand realities, and students must be free and responsible to decide what and how to learn (Perkins, 1991). Radical constructivism is problematic because while people continually create and change realities, they share many realities with common meanings and symbols such as history, language, religion, and values (Streibel, 1986).

(2) Moderate constructivists acknowledge reality that allows people to construct personal understandings of the world, but the world, but the world has physical and epistemological structures that constrain knowing (Cognition and Technology Group at Vanderbilt,1991). Communities impose constraints, but there are "sufficient degrees of freedom in the structure of physical and epistemological worlds to allow people to construct their own personal theories of their environments" (Cognition and Technology Group at Vanderbilt, 1991, p. 16). Moderate constructivists believe that "knowledge is a dialectical process the essence of which is that individuals have opportunities to test their constructed ideas on others, persuade others of the virtue of their thinking, and be persuaded" (Cognition and Technology Group at Vanderbilt, 1991, p. 16). Schools must create environments that capitalize on student ability to develop socially acceptable systems to explore new ideas. The goal of education is to "appreciate good rules, good theory, good science, good debate, etc" (Cognition and

Technology Group at Vanderbilt, 1991, p. 16).

(3) Rational constructivists acknowledge the dynamic nature of learning and the impossibility of predicting how students will learn. They recognize the weakness of being non-objective and anti-empirical, but maintain that people actively use new information with existing schema, interpret data, and organize this information into meaningful patterns. Knowledge transmitted by instructors may differ from what the learner constructs because the mind manages learning autonomously. Consequently, instructors should make instructional decisions when students cannot assume that responsibility. They should guide learners and provide concrete teaching, when necessary. Students, however, learn most effectively when they decide what and how to learn (Cognition and Technology Group at Vanderbilt, 1991; Winn, 1991).

Constructivism is inappropriate for entry-level learning when teachers establish objectives, curriculum, and methods. Objective instruction using a behavioral approach is necessary when learners lack sufficient background to construct knowledge. As learners acquire adequate backgrounds, teacher control can be gradually reduced. Students seldom take the entire responsibility for what and how to learn at the beginning of instruction (Perkins, 1991; Winn, 1991). In well-structured domains such as mathematics, physics, and chemistry, it is difficult for learners to create the required information (Spiro, Feltovich, Jacobson, & Coulson, 1991). For example, learners would have great difficulty calculating the area of a circle or proving why the sum of the three interior angles of a triangle is 180 degrees. Caseworker trainees cannot construct social service laws and regulations without reference materials. Specific counseling techniques can be taught, modeled, and practiced, but how a caseworker interacts with clients is a unique, personal construction.

Constructivism is appropriate for advanced level study in less structured domains such as social services, literature, and political science. In complex systems with vague content boundaries, people have legitimately different points of view. Different people inter-

preted the Vietnam War as an aggressive attack of the rights of a native population to protect the interests of multinational corporations or as an American obligation to defend people against the evils of communism (Jonassen, 1991b, p. 10). As students learn basic information, they become increasingly proficient in pursuing independent inquiries. When they reach this stage, they acquire the ability to apply and transfer knowledge to new situations (Spiro, Feltovich, Jacobson, & Coulson, 1991).

## Cognitive Apprenticeship

Throughout human history apprenticeship has a long legacy of learning, often with a parent or relative as a tutor. As society became more complex and specialized, formal training developed outside the family such as the guild system in medieval Europe. Apprenticeships in skilled crafts, such as chefs and electricians, continue in modern times, while universities provide high level professional training (Charness, Krampe, & Mayr, 1996). Unfortunately, professional level training in areas such as social work and education is often unrelated to the job-related demands that confront new workers.

Cognitive apprenticeship teaches the processes experts use to handle complex tasks. This learning-through-guided-experience process focuses on cognitive and metacognitive skills utilizing the external presentation of internal cognitive skills. This process is best achieved by observing how experts solve problems or by having experts describe how they accomplish certain tasks (Collins, Brown, Newman, 1989).

Cognitive apprenticeship includes the following techniques:

> *Modeling*: experts complete a task while learners observe and develop a conceptual model of the required processes. For example, a caseworker might model effective techniques for managing a hostile client and later, while observing a videotape of the incident, verbalize while her thought processes about why she acted in a particular way during the interview.

*Coaching*: experts observe learners complete tasks and offer hints, suggestions, and feedback.

*Articulation*: learners describe their knowledge, reasoning, and problem-solving processes. Reflection: learners compare their problem-solving processes with those of an expert.

*Exploration*: learners solve problems on their own. (Collins, Brown, Newman, 1989).

## ETE and Assessment Centers

ETE is implemented in a training setting that simulates the workplace. As Dubois (1993) noted, "Simulations emphasize a 'training-in-context' concept where the learning environment approximates the employee's workplace environment" (p. 193). Training in context means taking the time to build reality into your training design. It also means understanding how job pressures and organization climate affect the way work gets done. Most of all, it means that you have taken steps to help your trainees use new knowledge back on the job, where it will make a difference (Hendrickson, 1990, p. 70).

ETE utilizes the assessment center concept. As Reilly and Warech (1994) noted:

An assessment center is a comprehensive, standardized procedure in which multiple assessment techniques are used in combination to evaluate individuals for various organizational purposes. Although some assessment centers include paper-and-pencil tests and interviews, a special emphasis is placed on the use of situational exercises and job-related simulations. (p. 149)

In assessment centers, trained observers evaluate trainees as they demonstrate mastery of job-related skills. Trainees perform these behaviors in a simulated environment including group problem solving, role-play, oral presentations, in-basket, and videotaped exercises. These situational exercises elicit behaviors relevant to significant aspects of the target behavior (Reilly & Warech, 1994).

Research supports the concept of assessment centers. Tziner and Dolan (1982) found that assessment centers generally have higher validity for assessment ratings than paper-and pencil tests. Assessment centers also predict job performance. The AT&T Management Progress Study (Bray, 1964; Bray & Grant, 1966) compared ratings for 422 managers with their actual management performance eight years later. The predictive validity for the ratings was .44 for non-college-educated staff and .71 for college graduates. Reports from other companies also provided evidence regarding the predictive validity of assessment center ratings including IBM (Dodd, 1971), Standard Oil of Ohio (Finkle & Jones, 1970), Sears Roebuck (Bentz, 1969), and General Electric (Meyer, 1970).

Byham (1970) reported that 22 of 23 studies produced higher validity coefficients than alternative methods. Cohen, Moses, and Byham (1974) reported a median validity coefficient of .40 when promotion or other measures of job progress were used as indices, and a median validity of .63 when managerial potential was the criterion. Summaries by Huck (1977) and Klimoski and Strickland (1977) supported these findings. Meta-analyses by Schmitt, Gooding, Noe, and Kirsch (1984) and Hunter and Hunter (1984) reported comparable findings. In a meta-analysis of 50 assessment centers, Gaugler, Rosenthal, Thornton, and Bentson (1987) compared five criteria (job performance, potential, training performance, and career advancement) to four performance areas (promotion, early identification, selection, and research). The corrected mean across all validity subdivisions was .37 ranging from a low of .30 for promotion to a high of .48 for research. Since other factors, such as personality, health, and appearance, influence job success, the consistent, positive results from assessment center studies are particularly impressive.

The desirability of using realistic evaluation devices, such as those incorporated into assessment centers, is not new. In 1951 Lindquist observed that "the most important consideration is that the test question require the examinee to do the same things, however complex, that he is required to do in the criterion situations." (p. 154)

Assessing performance on job-related outcomes is an essential component of training. As Norceni and Shea (1993) noted:

> Outcomes are the ultimate criteria; they provide measures of the consequences of what is actually done in practice. . . to the public, they provide direct evidence that the practitioners are or are not achieving appropriate results. Outcomes assessment avoids many of the problems associated with traditional measures of competence because it is a measure of what happens in practice.

ETE is useful for assessing performance in CBT programs in which individual performance is judged against explicit standards that reflect expected outcomes based on competent job performance. If performance is based on work-place standards, then the logical way to determine if someone meets those standards is either on-the-job or in a setting similar to the job. Using work-related behavior samples to assess performance has been tested and found effective. Asher and Sciarrino (1974) demonstrated that tests based on "realistic" work samples related more strongly to later success than paper-based aptitude tests. Robertson and Kandola (1982) reported very high validity coefficients for work sample tests.

Since on-the-job assessment is often impractical and costly, a simulated environment allows trainers to scrutinize actual performance evidence that can be matched against specific standards of competence. ETE allows trainers to structure assessments that include conditions and contingencies impossible to assess in the workplace. In discussing the evaluation model used to assess the National Vocational Qualifications in the United Kingdom, Fletcher (1991) noted:

It would obviously not be practicable for an assessor to cause a deliberate breakdown of machinery (or indeed set fire to the building), simply to assess an individual's ability to cope. In this context, therefore, an assessor needs to be skilled in providing opportunities for supplementary assessment. This may involve a skills test, questioning of the individual, or allocating a new task or job. (p. 68)

## Constructive Testing

Constructed items test performance. They ask learners to supply, develop, demonstrate, create, or perform. Hambleton (1996) said that performance test have the following characteristics:

Assess what students know and can do with an emphasis on doing.

Use open-ended, constructed items to test higher level, cognitive skills.

Use direct assessment methods (e.g., writing sample to assess writing competence; counseling tapes to assess counseling skills).

Possess a high degree of realism.

May assess groups rather than individuals.

Conduct assessments over an extended period of time.

Include self-evaluation of projects or performances.

The setting or manner in which assessment is administered include different formats, as illustrated by the following list:

*Actors*: Professional actors perform using a script that presents job-relevant situations. Trainees interact with the actors to demonstrate significant practice behaviors. Observers use rating scales to evaluate trainee performance.

*Behavior modelin*g: After observing a demonstration of a model behavior, trainees practice the new behavior and receive feedback from trainers and/or peers. They can have multiple opportunities (either *in vivo* or videotaped) to practice until they ad-equately demonstrate the skill.

*Clinic*: Trainees describe job-related experiences and tell how they coped with situation. Their reported behaviors are self-evaluated by the presenting trainee and rated by trainers and participants.

*Case method*: Trainees critically examine an actual work situation that includes issues, circumstances, persons, actions, background information, documents and media surrounding the case. In a small or large group setting, trainees discuss issues and circum-stances surrounding the case and propose alternative ways to deal with the problem. Responses are graded using objective criteria.

*In-basket*: Trainees assume the role of a worker in a simulated agency and give written responses to letters and memos, make short audio-taped state-ments in response to telephone messages, and complete forms related to specific cases. Their answers are compared to poor, adequate, and good model responses.

*Laboratory*: In a setting that emulates workplace conditions, trainees are observed and evaluated as they demonstrate skills acquired during training.

*Portfolio*: Trainees assemble representative items (e.g., reports, documents, tapes) that illustrate their level of competence in relevant areas. This material is evaluated using objective criteria established by supervisors or experts.

*Role play*: Two or more trainees demonstrate skills in which they were trained. Variations such as multiple

role-play and role-reversal provide alternate ways to assess trainee competence.

*Clinical clients*: Medical schools test students by using clinical clients who display the symptoms of specific diseases and physical conditions. Clinical clients can provide trainees with realistic opportunities to demonstrate skills

*Skits*: Short, rehearsed, dramatic presentations involving two or more trainees who act from a prepared script to dramatize incidents that illustrate a situation.

*Video-tape performance*: Trainees complete video-tapes that demonstrate their competence in performing during an actual or simulated session. Trainer teams use prestated criteria to validate evaluate the videotapes based on a consensus among the judges.

The following example illustrates how ETE is integrated in a training activity. The activity, which includes three embedded evaluation items, is a 2-hour session designed to improve the level of empathic responding to clients. Figure 1 presents the 5-item test form.

*Activity 1*: Trainers complete a mini-lecture on using empathic understanding as a primary core facilitating dimension (Rogers, 1957; Carkhuff 1967; 1993). The mini-lecture builds on reading materials distributed as worksheets prior to training. The trainers model the technique by plying the roles of counselor and client in a simulated setting. (45 minutes)

*Activity 2*: Trainees watch short videotape set in the home of clients suspected of abusing a child. After the mother answers the counselor's initial statement, the counselor responds to clarify and reflect the emotional content of her comments. (15 minutes)

*Activity 3*: Trainees are given a scannable test form with a short vignette abstracted from the videotape.

*Activity 4*: Trainees form triads during which they rotate the roles of client, caseworker, and rater. Trainee-raters use the following 5-point scale to rate the level of empathy displayed by trainee-counselors and enter the rating on the answer sheet.

> *Empathy*: Responses to verbal and behavioral expressions of the client.
>
> > Level 5: expressed feelings more deeply than the client.
> >
> > Level 4: reflected deep, emotional feelings of group.
> >
> > Level 3: added to expressed feelings of group.
> >
> > Level 2: detracted from feelings of group.
> >
> > Level 1: detracted noticeably from feelings of group.

*Activity 5*: After each participant has played the role of a counselor, participants discuss specific behaviors that added to, or subtracted from, the level of displayed empathy.

*Activity 6*: Each triad reports back to the large group. (60 minutes)

*Activity 7*: After collecting answer sheets, trainers and trainees discuss the major conclusions to be drawn from the lesson. They discuss why client statements in items 1 through 4 received specific ratings and answer trainee questions. (45 minutes)

ETE procedures are described in specific, operational terms and include items, instruments, and scales used to assess trainee performance. Data collection demands are substantial since each trainee completes evaluation forms for every activity and records are maintained for each session and for every trainee. During a 4-day training

program, for example, eight or more forms might be collected for each trainee. Using machine-scorable assessment instruments facilitates scoring. Software packages, such as Teleform, which is published by Cardiff Software, produce attractive, scannable forms with the capability of scanning mark-sense and alphanumeric characters. These forms are optically scanned on a conventional flatbed scanner, preferably a high-speed duplex scanner. Forms can be processed from remote sites using a fax machine. Scanned data are automatically entered into a database, and database software, such as Access, can be used to produce different reports.

Appendix 1 shows a sample "Trainee Performance Report" that lists the trainee's name, county of employment, training dates, performance score, and mastery level. Instructions for remediation are provided for every unsatisfactory score. Other reports list each trainee in the group describing mastery status for each person for each content area, and pretest-posttest percentage improvement. Supervisors and staff development directors could use these diagnostic reports to prescribe individualized training programs to help staff master essential, job-related skills.

Appendix 2 presents a "Counselor Rating Scale" that includes two instruments for evaluating counseling competence. Appendix 3 contains a "Supervisor's OJT Checklist" that can be used to evaluate specific transfer-of-training effects at the agency level.

## Objective Test Items

Several types of objective items, such as completion and cloze procedure, are classified as constructed items. Cloze procedure is a diagnostic reading assessment technique. It deletes words from a passage according to a word-count formula or other criteria. Students respond by inserting words that construct meaning from the text.

Other objective items, such as true-false, matching, and multiple-choice, are usually classified as non-constructed. Multiple-choice items are widely used to test performance and achievement throughout society. They measure a wide range of content, performance, and psychological processes and are used for "high stakes" examinations to determine high school graduation, civil services employ-

ment, and admission to undergraduate, graduate, and professional programs. Objective tests are economical and efficient to administer and psychometrically superior to other item formats in regard to reliability. They test a wide range of content quickly and allow item analysis to diagnose specific student weaknesses.

Despite these advantages, many criticisms have been leveled against multiple-choice items. The major issue is that multiple-choice items do not test ability to develop and organize ideas and to develop coherent arguments or positions. In the worst case, the items test trivial knowledge and encourage guessing.

Properly constructed objective test items can assess a wide range of cognitive and attitudinal skills. Since expert performance requires a firm base of knowledge related to the topic of interest, objective tests can assess whether trainees have mastered the necessary prerequisites to demonstrate higher level skills. Objective items can also be used to assess higher level cognitive skills. For example, anecdotal items, either printed or videotaped, can be used as realistic stimuli for assessment. The following examples illustrate item formats that appropriate for this purpose.

## Illustration 1

Anecdotal item with a single best answer to assess process skills and higher order thinking.

### Instructions

Read the following excerpt and select the most empathic counselor response. An empathic response is one in which the counselor identifies underlying feelings and reflects content that complements the affective level of the client:

Client: Mary R., 27 years old, high school graduate, unemployed factory worker, fired seven months ago, is applying for Public Assistance. She is single with one child. This is the first counseling session.

"Don't you think that corporations are pretty heartless, even cruel? That is, they hire a person for their own purposes. They don't give a damn about you. Don't you think so?"

### Counselor Responses

a. You want to know what I think about corporations.

b. You feel strongly that corporations are all for themselves, and that a person who works for them just doesn't count.

c. In other words, you feel that corporations rather consistently violate the integrity of the people they employ.

d. You're depressed, maybe angry, because corporations don't have the proper concern and respect for their employees.

e. You feel that corporations have not hearts, and they hire people only for their own purposes without any concern about the people themselves.

## Illustration 2

Anecdotal item with weighted scoring to test higher level skills of evaluation and synthesis. This illustration describes different levels of empathy that are rated using a 7-point scale. This is an efficient method of testing that requires less testing time because each response is a separate item. Consequently, one anecdote represents five test questions.

### Instructions

Read the client statement for Nancy C. Then, read each counselor response and use the 7-point rating scale listed below to rate the level of empathy of each response.

1 = No attention to surface feelings

2 = Superficial awareness of surface feelings

3 = Minimal recognition of surface feelings

4 = Identifies surface feeling and emotion

5 = Accurately reflects surface feelings

6 = Reflects underlying emotions

7 = Enhances feelings and emotions

Client: Nancy C., 34 married, housewife, three children boys, 5 and 7, daughter 9. This is the second counseling session.

"My children are getting out of hand. They don't listen to me or my husband unless we threaten them. Who wants to threaten their children all the time? The oldest boy, Jimmy, was really well-behaved until last year and, then, suddenly he's a different kid. He's wild now, always yelling and screaming. Last week I caught him twisting his brother's arm. He really wanted to hurt him!

"I don't know where he gets it from. We're not a violent family. I could see it if I hit the kids all the time. But I don't. Oh, a swat now and then, but I never hurt them. My husband spank them hard sometimes, but he's not a mean person."

### Counselor Responses:

a. Raising children is a difficult job. It takes a lot of patience and understanding, but corporal punishment, from my point of view, is always wrong."

b. You're concerned and puzzled by your children's behavior, particularly Jimmy's.

c. You don't know what to do about their behavior."

d. Do you know whether anything is going on with the kids that has happened recently? Have you talked to them about this?

e. I don't want to tell you what to do, but maybe someone should talk to your husband about how he treats your children.

# Illustration 3

Anecdotal item as an indirect measure of values and attitudes. Rather than asking trainees to assess their own behavior, asks them to evaluate the behaviors of others. Ask trainees to select the best answer or to rate each distracter as demonstrated in the preceding illustration. The following example was selected from a test that assessed supervisory skills.

### Instructions

Read the client statement for Jim W. and the six actions a supervisor might take. Then, answer the following questions:

a. What is the most effective action for Jim's supervisor to take?

b. What is the least effective action for Jim's supervisor to take?

Client: Jim is a caseworker in an agency with a policy that prohibits employees from using the photocopy machine for personal business. Twice in the past month, after his supervisor observed him copy magazine articles unrelated to his job, he was warned that this was not allowed and that he should not do it again.

a. Fire him because he ignored agency policy.

b. Warn him that he will be terminated if he continues to ignore agency policy.

c. Charge him for the copies that he made.

d. Send him a letter describing what will happen if he continues this behavior.

e. Discuss the problem with your supervisor.

f. Ignore it because otherwise he is a good employee.

## Validity and Reliability

Validity is an overall judgment based on the theoretical rational and empirical evidence of the adequacy and appropriateness of inferences based on test scores. It is an inductive analysis of the adequacy of existing evidence for using and interpreting the test (Messick, 1988). Reliability is the extent to which test results are consistent, stable, and free of error variance that resulted from chance differences and different factors.

Constructed items have a distinct disadvantage compared with objective items in regard to reliability which is the extent to which test results are consistent, stable, and free of error variance. Error variance results from chance differences and is affected by different factors such as variations in examinee responses due to physiological or psychological conditions, testing conditions, scoring errors, and guessing.

It is difficult to develop valid constructive items. Hambleton (1996) cautioned:

> It should be emphasized that the new item formats, although more attractive in some respects than the common multiple-choice formats, are still required to meet psychometric standards of excellence. This point appears to have been lost by some performance assessment developers. For them, simply for performance assessments to be different in appearance from standard assessment procedures is sufficient to justify their use (p, 905).

Other experts have made similar conclusions in noting that performance measure often lack reproducibility and lack reliability (Mehrens, 1992). Tests that lack reliability are not viable politically, socially, or legally and should not be implemented on a large scale (Beck, 1991).

A full discussion of these issues is beyond the scope of this monograph. However, when creating constructed tests, developers must exert every effort to insure that validity and reliability are maximized. Procedures based on the judgment of expert panels have been developed and are appropriately used to develop criteria for validating constructed tests (Hambleton, 1996), while standard statistical methods can be used to determine reliability.

# Conclusion

This paper described a process called "Embedded Training Evaluation" which is appropriate for evaluating higher level cognitive skills and performance during training. The process uses test items that are blended into training activities which increases instructional time and results in improved trainee learning.

ETE provides opportunities for increased learning because trainers review correct answers immediately after tests are collected, and trainees can discuss their responses with the trainers and other trainees. Consequently, instructional time is increased because assessment, as an integral part of training, reinforces learning. C

Competence is enhanced because trainees practice and demonstrate skills in a relatively risk-free environment.

The ambiance of training is more professional since the conditions are comparable to those in which workers routinely perform their job responsibilities. Trainees use resource materials and ask questions of other trainees as they normally do during work. ETE eliminates the "high school examination" atmosphere of a formal posttest since trainees complete embedded items during training, and trainers collect multiple answer sheets for each trainee during training. An additional advantage is that ETE training materials are readily adapted to computer-based instruction.

# References

Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology*, 27, 519-533.

Beck, M. D. (1991, April). *Authentic assessment for large-scale accountability purposes: Balancing the rhetoric*. Paper presented as the annual meeting of the American educational Research Association, Chicago.

Bentz, V. J. (1969). Validity studies at Sears. In *Validities of assessment centers*, W.C. Byham, Chair. Symposium presented at the meeting of the American Psychological Association. Washington, D. C.

Berger, K. S., (1978). *The developing person*. New York, NY. Worth Publishers.

Billett, S. (1993). "What's in a setting? Learning in the workplace." A*ustralian Journal of Adult and Community Education*, *33*(1), 4-14.

Billett, S. (1994). Searching for authenticity. *Vocational aspects of education, 46*(2), 3-16.

Brainerd, C. J. (1978). *Piaget's Theory of Intelligence*. New Jersey: Prentice Hall, Inc.

Bray, D. W. (1964). The management progress study. *American Psychologist*, *19*, 419-429.

Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs*, *80*, 1-27.

Brooks, J. G. (1990). Teachers and students: Constructivists forging new connections. *Educational Leadership, 47*(5), 68-71.

Byham, W. C. (1970). Assessment center for spotting future managers. *Harvard Business Review, 48*, 150-160.

Carkhuff, R. R. (1967). *Beyond counseling and therapy*. New York: Holt, Rinehart and Winston.

Carkhuff, R. R. (1993). *The art of helping* (7th Ed.). Amherst, MS: Human Resource Development Press.

Cazden, C. (1981) Performance before competence: Assistance to child discourse in the zone of proximal development. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 3*, 5-8.

Charness, N., Krampe, R.., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: an international comparison of life-span chess skill acquisition. In K A. Ericsson (Ed.). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cognition and Technology Group at Vanderbilt. (1991, September). Some thoughts about constructivism and instructional design. *Educational Technology, 39*(9), 16-18.

Cohen, B. M., Moses, J. L., & Byham, W. C. (1974). *The validity of assessment centers: A literature review. Monograph II*. Pittsburgh, PA: Development Dimensions Press.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.

Dodd, W. C. (1971). Validity studies at IBM, In *Validity of assessment centers*. W. C. Byham, Chair, Symposium presented at the 79th annual convention of the American Psychological Association, Washington, D. C.

Dubois, D. D. (1993). *Competency-based performance improvement: A strategy for organizational change*. Amherst, MA: HRD Press.

Duffy, T. M., & Jonassen, D. H. (1991, May). Constructivism: New implications for instructional technology? *Educational Technology, 39(*5), 7-12.

Fabricius, W. V. (1983). Piaget's theory of knowledge: Its philosophical context. *Human Development*, *26*, 325-334.

Johnson, D & Johnson, R. (1987). *Learning together & alone: Cooperative, competitive, & individualistic learning*. Englewood Cliffs, NJ. Prentice-Hall.

Farmer, J. A., Jr., Buckmaster, A., & LeGrand, B. (1992).Cognitive apprenticeship. *New Directions in Adult and Continuing Education, 55*, 41-49.

Finkle, R. B., & Jones, W. S. (1970). *Assessing corporate talent: A key to management manpower*, New York: Wiley-Interscience.

Fletcher, S. (1991). *NVQs standards and competence: A practical guide for employers, managers and trainers*. London: Kogan Page.

Gagné, R. M. (1962a). The acquisition of knowledge. *Psychological Review, 69*, 355-365.

Gagné, R. M. (1962b). Military training and principles of learning. *American Psychologist, 17*, 263-276.

Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart & Winston.

Gagné, R. M. (1968). Learning hierarchies. *Educational Psychologist*, *6,* 1-9.

Gagné, R. M., Briggs L. J., & Wager W. W. (1992). *Principles of instructional design* (4th ed.). New York: Harcourt Brace Jovanovich.

Gagné, R. M., & Briggs, L. J., (1974). *Principles of instructional design*. New York: Holt, Rinehart and Winston.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, *72*, 493-511.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. Some questions. *American Psychologist*, *18*, 519-521.

Glaser, R. (1990). The reemergence of learning theory within instructional research. *American Psychologist, 45*(1), 29-39.

Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.). The nature of expertise. Mahwah, NJ: Lawrence Erlbaum Associates, XV-XXVIII.

Goodman, N. (1984). *Of mind and other matters*. Cambridge, MA: Harvard University Press.

Grabe, M. & Grabe C. (1998). *Integrating Technology for Meaningful Learning*. Houghton Mifflin.

Hambleton, R. K. (1996). Advances in assessment models, methods, and practices. In D. C. Berliner & R. C. Calfee (Eds.). *Handbook of educational psychology*. New York: Simon & Shuster Macmillan.

Hart-Landesberg, S., Braunger, J., & Reder, S. (1992). *Learning the ropes: The social construction of work-based learning*. Berkeley, CA: National Center for Research in Vocational Education, (ED 363 726).

Hendrickson, J. (1990, March). Training in context. *Training*, 65-70.

Huck, J. R. (1977). The research base. In J. L. Moses & W. C. Byham, (Eds.). *Applying the assessment center method*. New York: Pergamon Press.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98.

Klimoski, R. J., & Strickland, W. J. (1977). Assessment centers: Valid or merely prescient? *Personnel Psychology*, *30*, 353-63.

Jonassen, D. H. (1991a, September). Evaluating constructivistic learning. *Educational Technology*, *39*(9), 28-33.

Jonassen, D. H. (1991b, March). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology, 39*(3), 5-14.

Johnson, S. D., and Thomas, R. G. (1994). Implications of cognitive science for instructional design in technology education. *Journal of Technology Studies, 20* (1), 33-45.
Kliebard, H. (1986). *The struggle for the American curriculum 1983-1958.* Boston, MA: Routledge & Kegan Paul.

Lavatelli, C. (1973). *Piaget's Theory Applied to an Early Childhood Curriculum.* Boston: American Science and Engineering, Inc.

Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E.F. Lindquist (Ed.). *Educational measurement.* Washington, D.C.: American Council on Education.

London, C. (1988). A Piagetian constructivist perspective on curriculum development. *Reading Improvement*, *27*, 82-95.

Lynch, R. L. (1997). *Designing vocational and technical teacher education for the 21st century*. Columbus: ERIC Clearinghouse on Adult, Career, and Vocational Education.

Mehrens, W. W. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice, 11*(1), 3-9.

Messick, S. (1988). The once and future issues of validity: assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.). *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.

Molenda, M. (1991, September). A philosophical critique of the claims of "constructivism." *Educational Technology, 39*(9), 44-48.

Norcini, J. J., & Shea, J. A. (1993). Increasing pressures for recertification and relicensure. In L. Curry and J. E. Wergin (Eds.). *Educational professionals: Responding to new expectations for competence and accountability*. San Francisco: Jossey Bass.

Perkins, D. N. (1991, September). What constructivism demands of the learner. *Educational Technology*, *39*(9), 9-21.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.

Reilly, R. R., & Warech, M. A. (1994). The validity and fairness of alternatives to cognitive tests. In L. C. Wing & B. R. Gifford. (Eds.) *Policy issues in employment testing*. Boston: Kluwer Academic Publishers, 131-224.

Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, *55*, 171-183.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of consulting psychology*, *22*, 95-103.

Saunders, W. (1992). The constructivist perspective: Implications and teaching strategies for science. *School Science and Mathematics*, 92(3), 136-141.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, *37*, 407-422.

Sigel, I., & Cocking, R. (1977). *Cognitive Development from Childhood to Adolescence: A Constructivist Perspective.* New York: Holt, Rinehart and Winston.

Skinner, B. F. (1958). Teaching machines. *Science*, *128*, 969-977.

Skinner, B. F. (1958). *The technology of teaching.* Englewood Cliffs, NJ: Prentice-Hall.

Steffe, L. (1990). Overview of the action group A1: Early childhood years. In

L. Steffe and T. Wood (Ed.), *Transforming early childhood mathematics education: An international perspective.* Hillsdale, Lawrence Erlbaum.

Stevenson, J., (Ed.). (1994). *Cognition at work: The development of vocational expertise*. Leabrook, Australia: National Centre for Vocational Education Research, (ED 380 542).

Tharpe, R. G. and Gallimore, R. (1988). *Rousing minds to life*. Cambridge, MA. Cambridge University Press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplements, 2*, (Serial No. 8).

Thorndike, E. L. (1903). *Educational psychology*. New York: Teachers College.

Thorndike, E. L. (1918). The nature, purposes, and general methods of measurements of educational products. In *National Society for the Study of Education: 17th Yearbook, Part 2,* Bloomington, IL

Tziner, A., & Dolan, S. (1982). Validity of an assessment center for identifying future female officers in the military. *Journal of Applied Psychology*, *67*, 728-736.

Tobin, K. (Ed). (1993). *The Practice of constructivism in science education*. Washington DC. AAAS.

Tobin, K & Tippins, D. (1993) *Constructivism as a referent for teaching and learning.* In Tobin, K. (Ed.). *The Practice of constructivism in science education*. Washington DC. AAAS.

Roblyer, M. D., Edwards, J., & Havriluk, M. (1997) *Integrating Educational Technology into Teaching*, Merrill, Upper Saddle River, NJ.

Valsiner, J. (1993) Culture and human development: A co-constructivist perspective. In P. Van Geert & L. Moss (Eds.), *Annals of theoretical psychology*, Vol. X. New York: Plenum.

Verhave, T. (1959, March 26-27). Recent developments in the experimental analysis of behavior. In *Proceedings of the Eleventh Research Conference.* Chicago: University of Chicago.

Vygotsky, L. (1962). *Thought and language*. Cambridge, MA. MIT Press.

Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1981). The instrumental method in psychology. In J. Wertsch (Ed.). *Studies on the history of behavior: Ape, primitive, and child*. Hillsdale, NJ: Erlbaum.

Wheatley, G. H. (1991). Constructivist perspectives on science and mathematics learning. *Science Education* 75 (1), 9-21.

Wolf, A. (1995). *Competence-based assessment*. Buckingham, England: Open University Press.

Wiley, D. E., & Haertel, E. H. (1996). Extended assessment tasks: Purposes, definitions, scoring, and accuracy. In *Implementing performance assessment: Promises, problems, and challenges* (pp. 61-89). Mahwah, NJ: Lawrence Erlbaum Associates.

_____. (1991-1995). *Teleform 4 for Windows: User guide*. Carlsbad, CA: Cardiff Software, Inc.

# Appendix 1
# Sample Trainee Performance Report

## Trainee Performance Report

This report summarizes performance on embedded evaluation assessments administered during Child Welfare Core Permanency Planning training. Overall rating is the total score for these assessments, and mastery requires a minimum score of 75%.

Specific instructions for supplementary training are provided when test performance did not reach the required competence level for a content area. Please note that test performance is only one measure of a worker's ability, and that test performance is contingent upon a number of factors.

We recommend that supervisors use this information as a guide to help the caseworker master areas in which performance did not meet expectations. Please feel free to contact us at 1-800-796-7600 if you have any questions.

*Name*: Smith, John  *County*: Albany  *Performance Score*: 81
*SS #*:117-23-1379  *Training Dates*: 1/11-28/99  *Mastery*: Yes

Summary of participant performance: Mr. Smith's performance score of 81% exceeded the minimum mastery level of 75%. His score was 4 points below the mean of the training group.

Satisfactory Performance:

Stated goals and objectives.

Demonstrated genuineness with client during counseling simulation.

Identified factors related to identifying child safety.

Identified factors related to identifying level of risk.

Described appropriate family support systems.

Defined legal concepts relevant for child safety and risk.

Unsatisfactory Performance

Demonstrated empathic understanding for client during counseling simulation.

Review related material in workbook on pages 34-47.

Read articles by Robert Carkhuff and Carl Rogers which were included in the workbook.

Complete exercise 3 on page 37 in workbook.

Demonstrated positive regard for client during counseling simulation.

Review related material in workbook on pages 48-61.

Read article by Carl Rogers, which was included in the workbook.

Complete exercise 5 on page 52 in workbook.

# Appendix 2
# Counselor Rating Scale

## Counseling Rating Scale

### Part A:

**Instructions**: Please evaluate the counselor's performance during the sessions you recently completed. Check the appropriate level by darkening the corresponding box.

**Empathy**: Responses to verbal and behavioral expressions of the client.

- ❏ Level 5: expressed feelings more deeply than the client.
- ❏ Level 4: reflected deep, emotional feelings of group.
- ❏ Level 3: added to expressed feelings of group.
- ❏ Level 2: detracted from feelings of group.
- ❏ Level 1: detracted noticeably from feelings of group.

**Respect**: Counselor respect for client feelings, experience, and potential.

- ❏ Level 5: deepest respect and concern.
- ❏ Level 4: deep respect and concern.
- ❏ Level 3: respect and concern.
- ❏ Level 2: little respect and concern.
- ❏ Level 1: lack of respect or disrespect.

**Genuineness**: Relationship between verbal and physical expressions of the counselor.

- ❏ Level 5: perfect match.
- ❏ Level 4: close match; not perfect.
- ❏ Level 3: no obvious discrepancies.

❏ Level 2: somewhat unrelated; mildly negative results.

❏ Level 1: clearly unrelated; negative; destructive.

**Concreteness**: Client discussed personally relevant material in specific, concrete terms.

❏ Level 5: discussed specific feelings fluently and completely.

❏ Level 4: discussed concrete, specific problems.

❏ Level 3: required counselor facilitation.

❏ Level 2: discussed personal material in vague, abstract terms.

❏ Level 1: required counselor direction or dealt with vague generalities.

**Self-exploration**: Client introduced personally relevant material

❏ Level 5: actively engaged in inward probing.

❏ Level 4: discussed personal material with emotion.

❏ Level 3: introduced personal material without emotion.

❏ Level 2: responded unemotionally to personal material introduced by counselor.

❏ Level 1: discussed personal material unwillingly.

**Part B**:

**Instructions**: Please rate the counselor's performance during the session you recently completed. Use the following scale to rate each behavior.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Strongly disagree | Disagree | Not sure | Agree | Strongly Agree |

| During the group counseling sessions I . . . | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| . . . was concise and to the point. | | | | | |
| . . . ended on time, but not abruptly. | | | | | |
| . . . responded appropriately using meaningful questions. | | | | | |
| . . . made interpretations that were concise and to the point. | | | | | |
| . . . responded non-judgmentally. | | | | | |
| . . . responded appropriately neither interrupting nor waiting too long. | | | | | |
| . . . made responses consistent with what the client said. | | | | | |
| . . . earned the respect of the client. | | | | | |
| . . . possessed effective counseling skills. | | | | | |
| . . . encouraged participation of the client. | | | | | |
| . . . used natural, spontaneous expressions. | | | | | |
| . . . understood nonverbal client behaviors. | | | | | |
| . . . reflected important client concerns. | | | | | |
| . . . assessed client problems accurately. | | | | | |
| . . . confronted the client appropriately. | | | | | |
| . . . dealt effectively with crisis situations (e.g., suicide, abuse). | | | | | |
| . . . worked effectively with unmotivated the client. | | | | | |
| . . . worked effectively with nonverbal the client. | | | | | |
| . . . worked effectively with different social and ethnic groups. | | | | | |
| . . . helped the client define specific problems. | | | | | |
| . . . expressed thoughts and feelings clearly. | | | | | |
| . . . was genuinely relaxed and comfortable. | | | | | |
| . . . displayed spontaneous, relevant behavior. | | | | | |
| . . . dealt with content and feeling during sessions. | | | | | |
| . . . displayed self-confidence in session. | | | | | |
| . . . demonstrated interest in client concerns. | | | | | |

## References

Benshoff, J. M., & Thomas, W. P. (1992). A new look at the counselor evaluation rating scale. *Counselor Education and Supervision, 32,* 12-23.

Carkhuff, R. R., & Berenson, B. G. (1967). *Beyond counseling and therapy. New York: Holt, Rinehart and Winston.*

Larson, L. M., Suzuki, L. A., Gillespie, K. N., Potenza, M. T., Bechtel, M. A., Toulouse, A. L. (1992). Development and validation of the counseling self-estimate inventory. *Journal of Counseling Psychology, 39*(1), 105-120.

# Appendix 3
# Supervisor's OJT Checklist

### Instructions

New York State Permanency Planning Common Core training helps workers meet their developmental needs and the challenges that will face them in their new jobs. Although evaluation of worker skills is an integral part of residential training, the true test of learning is ability to apply skills on the job. Supervisors can help new workers develop these skills through monitoring, coaching, and feedback.

The scale includes three Core Conditions (Respect, Empathy, and Genuineness) and five Interpersonal Helping Skills (Attending, Questions, Reflections, Concreteness, and Summarization). Please use the following scale to rate how effectively the worker used each core condition and skill on this scale.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

Indicate your answer by drawing a circle around the number that best reflects the worker's level of performance. Add additional comments in the space provided.

Thank you for completing this scale. We are confident that you and the worker will consider it time well spent in improving child welfare practice.

# Rating Scale
# Core Conditions

Rate each core condition and interpersonal helping skill globally. In other words, assess the extent to which the caseworker demonstrates all, or most of, the characteristics that describe the core condition.

### Empathy

Made responses that are interchangeable with client expressions.

Communicated understanding of, and compassion for, client experience.

Recognized non-verbal cues.

Showed a desire to comprehend.

Discussed what is important to the client.

Referred to client feelings.

Reflected implicit messages.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Genuineness

Was honest and open.

Matched verbal and nonverbal behaviors.

Acted spontaneously.

Acted non-defensively.

Used self-disclosure.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Respect

Exhibited respect for client feelings, experience, and potential.

Demonstrated commitment.

Communicated warmth.

Valued clients simply because they are human beings.

Reinforced client strengths.

Suspended critical judgement.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

## Interpersonal Helping Skills

### Attending

Recognized level of congruence of client verbal and non-verbal behavior.

Used minimal encouragers.

Suspended critical judgement.

Conveyed respect for, acceptance of, and interest in the client.

Created a comfortable environment by removing physical barriers and minimizing distractions.

Used nonverbal behaviors (e.g., eye contact, posture, gestures, facial expressions, voice quality, and spacing of words).

Attended to client non-verbal and para-verbal behaviors.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Questioning

Encouraged the client to share information that deepens worker and client understanding.

Used open-ended questions that invite others to talk.

Asked questions in a warm, concerned manner.

Maintained effective eye contact, voice tone, and facial expressions.

Used closed questions to focus on specific information.

Used indirect questions that imply, but do not directly ask a question.

Used scaling questions that ask clients to rank-order or rate something.

Asked circular questions expand the field of inquiry to include the feedback the client is getting from others.

Used solution-based questions to build self-concept.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Reflection

Communicated understanding of client feelings, behaviors, values, beliefs, needs, and self-concept.

Listened to words and feeling in the message.

Observed non-verbal cues

Matched feeling in the client's message

Represented the content in the client's message

Matched the client's sensory words

Considered non-verbal cues.

Focused attention on the client's message.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Concreteness

Verified client perceptions.

Helped client be specific about feelings, experiences, and behavior.

Clarified vague or unfamiliar terms.

Explored reasons for conclusions.

Helped clients personalize statements.

Elicited specific feelings.

Elicited details.

Focused on the here-and-now.

Modeled concreteness in responding to clients.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Summarization

Used brief summarizations.

Structured the interview.

Focused discussion.

Made transitions.

Synthesized facts and feelings communicated by the client..

Summarized during breaks in interview or when clients deviate from major issues.

Conveyed acceptance of client perspectives and experience.

Checked for accuracy with the client.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

*Open-ended comments:* Please make additional comments describing the worker's performance.

# Appendix 3
# Supervisor's OJT Checklist – Module 2

**Instructions**

New York State Permanency Planning Common Core training helps workers meet their developmental needs and the challenges that will face them in their new jobs. Although evaluation of worker skills is an integral part of residential training, the true test of learning is ability to apply skills on the job. Supervisors can help new workers develop these skills through monitoring, coaching, and feedback.

The scale is based on material covered during Module 2. It includes worker competencies for four family characteristics: Child and Adult Development, Strength-Based Skills, Promoting Conditions Necessary for Change and Family-Focused Intervention.

Please use the following scale to rate how effectively the worker used these competencies.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

Indicate your answer by drawing a circle around the number that best reflects the worker's level of performance. Add additional comments in the space provided.

Thank you for completing this scale. We are confident that you and the worker will consider it time well spent toward their ability to face the challenges of child welfare practice.

# Rating Scale

## Child and Adult Development

Recognize stages of development for various family members.

Gather information on tasks and needs of particular stages.

Interact with family members according to information gathered.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

## Strength-Based Skills

Used exception-finding questions.

Used lottery questions.

Used scaling questions.

Used coping questions.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

## Promoting Conditions Necessary for Change

Assessed present discomfort.

Assessed emotional security.

Assessed internalization of responsibility.

Assessed sense of efficacy.

Assessed preferred alternative future.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Assessing Behavior and Underlying Conditions

Explored needs underlying the behavior of family members

Explored factors and conditions in and between the family and environment that might be influence behavior.

Explored the parent's beliefs and values regarding care of children.

Gathered information regarding parental self-concept and its impact on their behavior.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

### Family Focused Intervention

Encouraged interaction among family members.

Sanctioned expression of feelings between family members.

Acknowledged the impact of members' actions on others.

Used gatekeeping to block or elicit family member contributions to the conversation.

Gathered information regarding system components (i.e., Who has the power? Who makes the rules? What roles do family members fill? What subsystems are present?).

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very ineffective | Moderately ineffective | Mildly ineffective | Mildly effective | Moderately effective | Very effective | Not applicable |

*Open-ended comments:* Please make additional comments describing the worker's performance.