

ASSESSMENT METHODS

(Information Synthesized and Organized by Nancy Sinclair, MBA, RN, Assessment & Learning, UNM SOM)

“If you don’t know to which port you are sailing, no wind is strong enough...” Lucia Annaeus Seneca

“The quality of the assessment can only be considered in the context of the purpose.” Jay Parkes, Ph.D.

Creating Meaningful Assessment (Jay Parkes, PhD, Assoc. Professor Educational Psychology, University of New Mexico, 2004)

1. Determine objectives
2. Determine content
3. Determine how students should demonstrate that they have met the objective(s)
4. Make a test blueprint
5. Plan and deliver your instruction
6. Develop the assessment components

FINK'S TAXONOMY (Fink, *Creating Significant Learning Experiences*, 2003)

Foundational Knowledge	Application	Integration	Human Dimension	Caring	Learning to Learn
<ul style="list-style-type: none"> ■ Understanding ■ Remembering 	<ul style="list-style-type: none"> ■ Thinking: <ul style="list-style-type: none"> ○ Critical ○ Practical ○ Creative ■ Managing Complex Projects ■ Performance Skills 	<ul style="list-style-type: none"> ■ Interdisciplinary Learning ■ Learning Communities ■ Learning & Living/Working 	<ul style="list-style-type: none"> ■ Interpersonal ■ Self Authorship ■ Leadership ■ Ethics ■ Character ■ Multicultural Education ■ Teams ■ Citizenship ■ Environmental Ethics 	<ul style="list-style-type: none"> ■ Wanting to Be a Good Learner ■ Developing Keen Interest ■ Responsibility for Self 	<ul style="list-style-type: none"> ■ Becoming a Better Learner ■ How to Inquire and Construct ■ How to Pursue Self-Directed Learning

Assessment Method	Description	Fink's Taxonomy	Advantages	Cautions/Disadvantages
Computer Based Testing	Use of computers or electronic media as an assessment modality	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Integration 	<ul style="list-style-type: none"> ■ Available any time ■ Can be programmed to respond individually to student's responses ■ Capable of assessing a large number of learners ■ Allows student self-assessment with immediate feedback 	<ul style="list-style-type: none"> ■ Comparisons between computer simulation and real life present validity concerns
Direct Observation	Observation by a teacher of learner performance	<ul style="list-style-type: none"> ■ Integration ■ Caring ■ Human Dimension ■ Learning to Learn 	<ul style="list-style-type: none"> ■ Validity of feedback ■ Opportunity for formative feedback 	<ul style="list-style-type: none"> ■ Need multiple observations to establish reliability for scoring purposes
Essay	Open-ended response to question (stem or vignette) where students are required to write a short or long paragraph to answer the question	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Caring ■ Human Dimension 	<ul style="list-style-type: none"> ■ Encourages freedom of expression and creativity ■ Emphasizes depth and scope 	<ul style="list-style-type: none"> ■ Time consuming to grade (develop scoring rubric and training to score) ■ Requires well constructed test-questions
Global Rating Scale	Allows for scoring on a continuum. Each point of the scale is anchored to an observable behavior	<ul style="list-style-type: none"> ■ Application ■ Integration ■ Human Dimension 	<ul style="list-style-type: none"> ■ Measures complex skills ■ Anchors abstract skills to behavioral objectives ■ Low cost ■ Can be used with direct, indirect and memory of learner performance ■ Provides behaviorally specific feedback ■ Reflects developmental performance on a continuum 	<ul style="list-style-type: none"> ■ Inter-rater reliability ■ Subject to the same list of rater errors (listed under the Peer Evaluation Method) ■ Requires extensive rater training to promote reliability
In-Basket	Collection of materials that might be found in an in-basket in the domain being assessed; student's task is to manage each task appropriately within a specified time limit	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Integration ■ Human Dimension 	<ul style="list-style-type: none"> ■ Realistic situation ■ Standardized ■ Useful in teaching and training ■ Can be self-administered and self assessed 	<ul style="list-style-type: none"> ■ Labor intensive to design ■ Requires expert rater to establish scores

Assessment Method	Description	Fink's Taxonomy	Advantages	Cautions/Disadvantages
Modified Essay Question	Description of an evolving situation with questions asked of the learner at frequent junctions in either MCQ or essay format. Once the question is answered, more information is revealed.	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Integration 	<ul style="list-style-type: none"> ■ High validity for measuring understanding of complex situations 	<ul style="list-style-type: none"> ■ Must be hand-graded ■ May involve complex scoring scheme ■ Administration challenges with student not being allowed to look ahead to the next section until they complete the question.
Multiple Choice Question (MCQ) (includes one best answer and extended matching)	Test question where a stem is associated with several alternative answers, only one of which is (most) correct.	<ul style="list-style-type: none"> ■ Foundational Knowledge Challenging to Construct: <ul style="list-style-type: none"> ■ Application ■ Integration 	<ul style="list-style-type: none"> ■ Ease / speed of scoring ■ Objective scoring ■ High reliability if well constructed ■ Covers a large body of knowledge at once ■ Allows wide sampling of content areas ■ Calibrates easily for test equivalence 	<ul style="list-style-type: none"> ■ Reinforces “right/wrong” concepts in knowledge ■ Potential for “correct answer” by guessing ■ Unrealistic ■ No clues about thought process of learner ■ Encourages adapting learning to test format
Peer Evaluation	Review of an aspect of performance based on evaluation tools, perceptions and/or observations of a member of the person's peer group	<ul style="list-style-type: none"> ■ Application ■ Human Dimension ■ Caring ■ Learning to Learn 	<ul style="list-style-type: none"> ■ May identify strengths and weaknesses earlier than with other methods ■ Allows learners to develop as teachers and practice constructive feedback skills 	<ul style="list-style-type: none"> ■ Inconsistent assessment abilities among students ■ Error tendencies: <ul style="list-style-type: none"> Leniency Halo Effect Central Tendency Logical Self-comparison Proximity
Practical or Performance	Requires application and demonstration of knowledge. May include stations where learner demonstrates a particular skill	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Integration 	<ul style="list-style-type: none"> ■ Efficient ■ Objectively scored ■ Standardized stations ■ Creativity in development ■ Engaging for students and faculty ■ Allows for varied materials and settings ■ Opportunity to tailor specifically to learning objectives 	<ul style="list-style-type: none"> ■ Time intensive for design phase ■ Requires many stations to achieve reliability ■ Identifying a “gold standard” for scoring ■ May be subjective
Oral	Examiner asks the learner a series of questions to which he/she must respond verbally.	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Integration ■ Caring ■ Human Dimension 	<ul style="list-style-type: none"> ■ Flexible ■ Measures unique abilities (such as thinking on one's feet or following a line of heard reasoning) ■ Encourages teaching-learning exchanges during the exam ■ Allows for immediate feedback ■ Motivating 	<ul style="list-style-type: none"> ■ May have less reliability due to inconsistencies within or between the examiner(s) ■ The learner may not be able to demonstrate his/her knowledge in this context because of anxiety or other factors ■ Training the examiner to administer and score the exam consistently

Assessment Method	Description	Fink's Taxonomy	Advantages	Cautions/Disadvantages
Self-Assessment	Assessing one's own performance based on established criteria; also, assessing one's own attitudes, beliefs, values, opinions by completing an inventory type of instrument	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Human Dimension ■ Caring ■ Learning to Learn 	<ul style="list-style-type: none"> ■ Helps learners develop metacognitive skills ■ Provides student immediate feedback ■ Is time efficient ■ Using an audience response system for this can allow students to see their answers in comparison to others in an anonymous way ■ Can be combined with peer and teacher assessment, allowing student to see self-assessment in comparison to colleagues and an expert (such as with the use of <i>Calibrated Peer Review</i>) 	<ul style="list-style-type: none"> ■ If for grade, honor code considerations ■ May be most effective when used in conjunction with other methods and for self-discovery purposes
Triple Jump	<p>Three-step examination to evaluate performance of reasoning and application, self-directed learning and self-assessment</p> <ol style="list-style-type: none"> 1. Question is assigned to student 2. Student is given time to write an essay in response 3. Student presents response (synthesis of the problem) 	<ul style="list-style-type: none"> ■ Foundational Knowledge ■ Application ■ Human Dimension ■ Caring ■ Learning to Learn 	<ul style="list-style-type: none"> ■ Direct feedback to student ■ Provides opportunity for self-reflection in the context of a teacher ■ 360 degree assessment between student and teacher 	<ul style="list-style-type: none"> ■ Reliability may be low depending on <ul style="list-style-type: none"> The examiner Type of oral exam, material & context Candidate's personality Emotional reaction to exam Training of faculty

References: CBT:(Norman & Prinvin, 1985), Direct Observation: (Andrew, 1977; Wakefield, 1985), Essay Question: (Derova, 1979; Neufeld, 1985), In-basket Exercise (Andes, 1977; Hakistan, Woolsey, & Schroeder, 1986), Modified Essay Question: (Knox, 1980), MCQ: (Friedman et al., 1987; Norcini et al., 1985; Scriven, 1991; Zelezkik et al., 1990), Practical: (Harden et al., 1975; Newble & Swanson, 1988; Frye et al., 1989; van der Vleuten & Swanson, 1990; Petrusa et al., 1990 Miller, 1990, Lunenfeld et al., 1991), Oral Exam: (Mizzin & Hart, 1985), Peer Evaluation (Carline, Wenrich & Ramsey, 1989; Helfer, 1972), Global Rating Scale: (Keynan, Friedman & Benbassat, 1985), Triple Jump: (Neufeld & Norman, 1985; Painvin et al., 1979)