NATURE, PURPOSES & PRACTICES OF ASSESSMENT

Dominador D. Mangao
Research & Development Division
SEAMEO RECSAM, Penang, Malaysia
4 April 2019

# *"To fail to assess is to fail to teach…"*
(Best, 1992)

Constructivist Manifesto

"If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly"
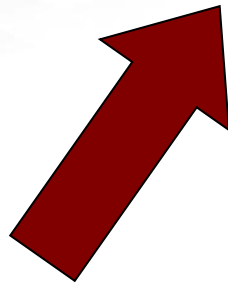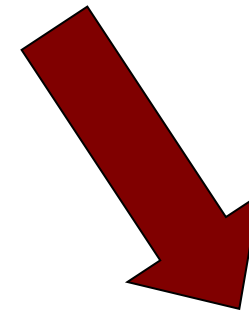
(Ausubel, 1978)

STUDENT ASSESSMENT CYCLE

TEACHING

ASSESSMENT

LEARNING

Should

Improve

Stimulates

Facilitates

Subjected

To
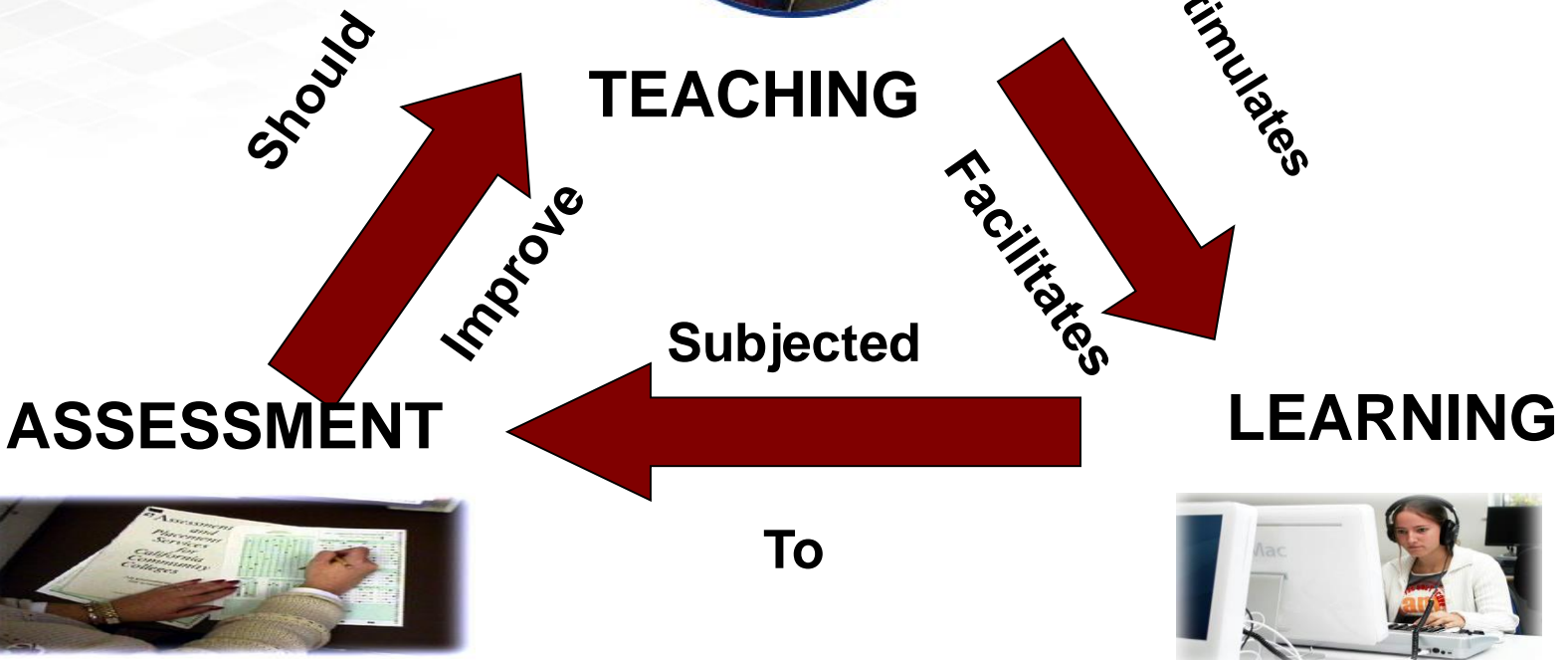
CURRICULUM ALIGNMENT

Learning Goals/ Standards

Assessment

Teaching/Instruction

Principle of Congruence

1. How does the teacher take a **standard/learning goal** and use it to **design an assessment** that provides valid and sufficient evidence that this goal has been achieved?

2. How does a teacher use this **assessment** to guide his or her **selection of learning experiences** that enable students to demonstrate that they have attained the learning goals?

To have standards, instruction, and assessment aligned and consistent, teachers can ask themselves three questions:

1. What do my students need to know and be able to do? (Standards)

2. How will I know whether my students meet the standard? (Assessment)

3. What learning opportunities will I provide for students to meet the standard? (Instruction)

Instructional Goals and Objectives: Foundation for Assessment

- Instructional **goals and objectives** serve as guides for both teaching and learning, communicate the intent of the **instruction** to others, and provide guidelines for **assessing student learning.**

- What types of learning outcomes do you expect from your teaching? Knowledge? Understanding? Applications? Thinking skills? Performance skills? Attitudes?

- Sound assessment requires relating the assessment procedures as directly as possible to intended learning outcomes.

Instructional Objectives as Learning Outcomes

- Instructional goals and objectives are sometimes stated in terms of **actions** to be taken.

- A better way to state objectives is in terms of what we expect students **to know** and be **able to do** at the end of instruction.

- Well- stated outcomes make clear the types of student performance we are willing to accept as evidence that the instruction has been successful.

- Example: Demonstrate to students how to use the microscope

- Demonstrate skill in using microscope to identify features of a cell

Assessment, Test, and Measurement

- **Assessment** is a general term that includes the full range of procedures used to gain information about student learning (observations, ratings of performances or projects, paper-and-pencil tests) and the formation of value judgments concerning learning progress.

- Assessment may include both quantitative descriptions (measurement) and qualitative descriptions (non-measurement) of stude

- It answers the question *"How well does the individual perform?"*

- A **test** is a particular type of assessment that typically consists of a **set of questions** administered during a fixed period of time under reasonably comparable conditions for all student.

- **Test**: An instrument or systematic procedure for measuring a sample of behavior by posing a set of questions in a uniform manner. Because a test is a form of assessment, tests also answer the *question "how well does the individual perform – either in comparison with others or in comparison with a domain of performance tasks?"*

- **Measurement** is the **assigning of numbers** to the results of a test or other type of assessment according to a specific rule (e.g. counting correct answers or awarding points for particular aspects of an essay).

- **Measurement**: The process of obtaining a **numerical description** of the degree to which an individual possess a particular characteristic. *Measurement answers the question "How much?"*

- It is limited to **quantitative descriptions** of students; that is, the results of measurement are always expressed in numbers.

- It does **not include qualitative descriptions**, nor does it imply judgments concerning the worth or value of the obtained results.

Assessment

- "assess" French word, "assidere", "to sit beside".

- process :

  i.   by which two or more people
     sit together to confer

  ii. to determine "How we are doing?
     How can we do better?"

Dominador D. Mangao, Science Specialist
SEAMEO RECSAM, Penang, Malaysia

Assessment

- non-threatening tone and affirming
- suggests a partnership based on mutual trust and understanding
- reminds a positive association between assessment and the process of teaching and learning
- broadest sense: concerned with children's progress and achievement

Assessment

- a way of continually facilitating students' :
- Interest
- Engagement
- Confidence
- Learning competencies (Carole Cooper, 1998)

- is the process of …
- collecting ,synthesizing ,interpreting information, aid in decision-making.

•Assessment is not just a means of gathering data;

•it is a means to reflect on one's experiences so one can learn and grow from them.

Classroom Assessment

is the process of …

- gathering ,recording ,interpreting, using and communicating information about a child's progress and achievement during the development of knowledge, concepts, skills and attitudes

- involves much more than testing

- ongoing process that encompasses formal and informal activities to monitor and improve teaching and learning

Dominador D. Mangao, Science Specialist
SEAMEO RECSAM, Penang, Malaysia

Changing Assessment Philosophy:

Necessary Mind shifts About Assessment

- All students can and are learning.
- The overall goal of assessment is self-assessment.
- We need to use multiple forms of assessment.
- We need to use multiple assessors (self, teachers, peers, parents).
- We need to look for success and movement toward desired outcomes or learning results.

Changing Assessment Philosophy:

Necessary Mind shifts About Assessment
- Human judgment can be reliable and valid.
- The assessment is based on measurable, precise criteria.
- The criteria is NO SECRET to the learners.
- We need to assess the PROCESS as well as the PRODUCT,

   (i.e. assess meta-cognition (how one thinks) and meta-learning (how one learns).

Changing Assessment Philosophy

- **Necessary Mind shifts About Assessment**

- Assessment is ON-GOING and continuous.

- Assessment is NOT FINITE; it should not be limiting or set limits on student's learning.

- Assessment results need to GUIDE INSTRUCTION.

- The assessment, itself, is a LEARNING EXPERIENCE

Assessment: The Big Picture

The trend in science education reform is toward assessments that:

- Are **embedded** within instructional materials
- Use a **variety of methods** to assess student progress
- Emphasize teacher observation, questioning, feedback and judgment
- Provide methods for getting at the **reasons behind student answers**

Purposes of Assessment

- To assist student learning
- To identify or diagnose students' strengths and weaknesses
- To assess the effectiveness of a particular instructional strategy
- To assess and improve the effectiveness of curriculum programs
- To assess and improve teaching effectiveness
- To provide data that assist in decision making
- To communicate with and involve parents (Kellough et al (1999)

The purpose of assessment in education

Three broad categories: summative, formative and quality assurance.

The main purposes of assessment are:

a. concerned respectively with the support of learning,

b. with certification (i.e., with reporting the achievements of individuals) , and

c. with satisfying demands for public accountability (Black and William,2007) .

Competencies to be Assessed

- **Knowledge – conceptual, declarative**:

- the content standards and key understanding of a topic or content/learning area

- **Knowledge – skills, procedural**: Understanding and use of skills, the steps or parts of concept, such as reading a map or computing an answer

Competencies to be Assessed

- **Learning Processes:**

- Critical, creative, and reflective thinking skills; group and interactive learning skills; how one takes on a task and works through it to gain insights and learning

Competencies to be Assessed

- **Attitudes, Dispositions, Habits of Mind, Work Habits:**

- Values toward learning; positive attitudes toward working, thinking, or learning such as curiosity, independence, managing time, persistence

- **Learning Products:**

- demonstrate conceptual, procedural, learning processes and attitudes and work habits of learning

Matching Best Methods to Student Outcomes

Conceptual, declarative Knowledge	Performances Interviews Constructed Response Tests
Procedural Knowledge	Observation + Interviews Portfolios
Learning Processes /Products	Performances Portfolios
Attitudes/Work Habits	Interviews Observations Work Samples

Matching Assessment Methods to Student Competencies/Outcomes

Competency	Written Tests	Observation/ Interview	Portfolios/Work samples	Performances/ Exhibitions
Conceptual Knowledge	Essays can show understanding, not ability to use it	Good for narrow range of knowing	Not good	Strength: Use of reference material; shows relationships of ideas in product or performance
Procedural Knowledge	Can describe steps, does not show they can use it	Strength: for demonstrating steps and skills	Good for showing progress and use of steps	Good for showing complex use
Learning Processes, Products	Not good, Essays could describe steps or how do, may not be able to do	Good if student talks out loud; can probe their thinking, steps in making product	Strength if assesses with portfolio	Strength for products and can assess the carrying out of processes and steps
Attitudes/Work Habits	Questionnaires or open-ended sentences in essay are good	Can discuss feelings, see how student approaches subject	If student writes about it; or can infer from artifacts	Can only infer from observation or product or presentation

Matching Best Methods to Student Outcomes

Assessing Processes	Interviews Observations, written data Learning logs and Journals Oral/Written Self-Evaluation Debriefing Interviews about projects, products, etc. Behavioral Checklists Rating Scales, Feedback Form Student self-Talk Quizzes, Progress Reports Process-folios

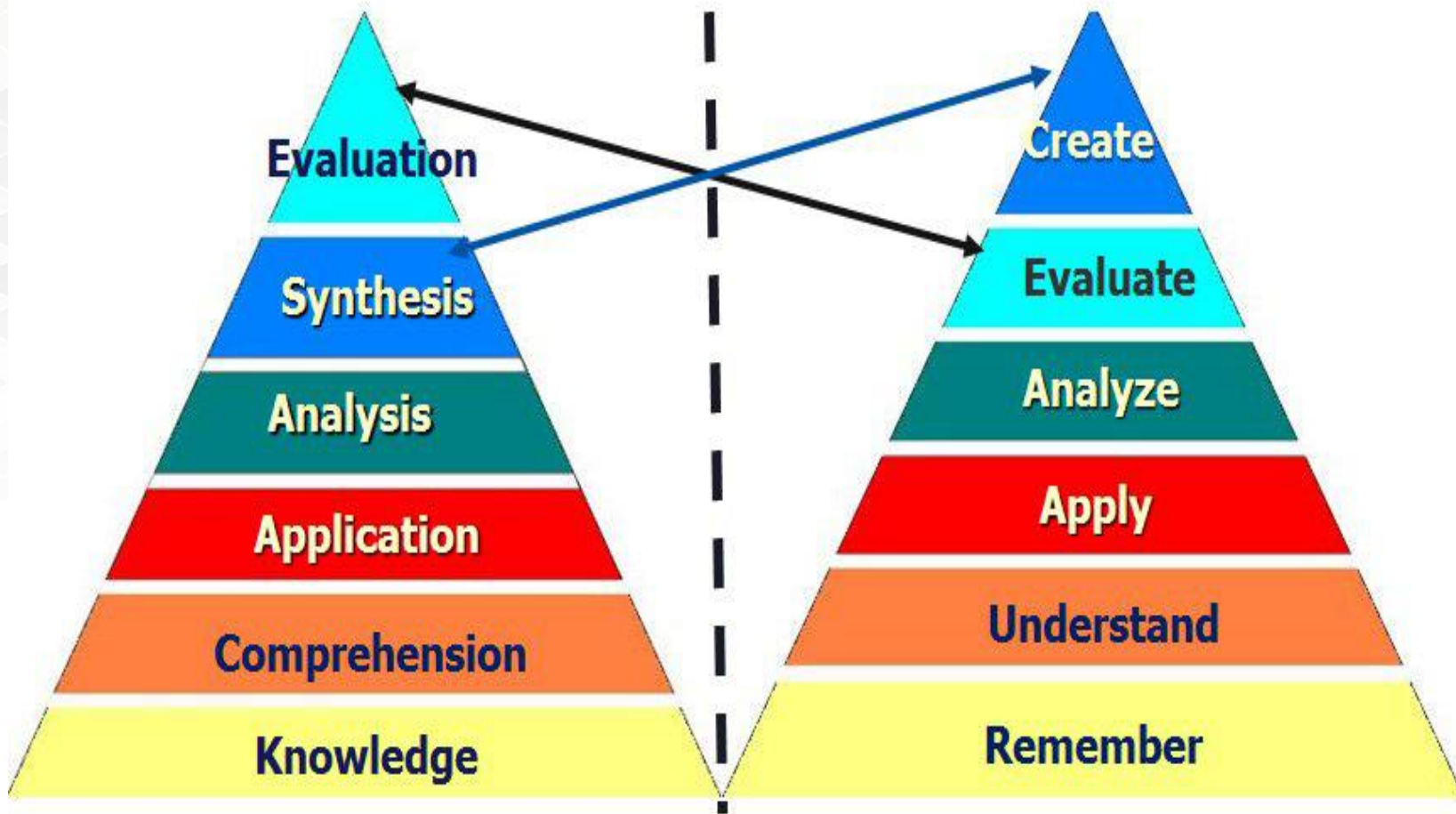# Matching Best Methods to Student Outcomes

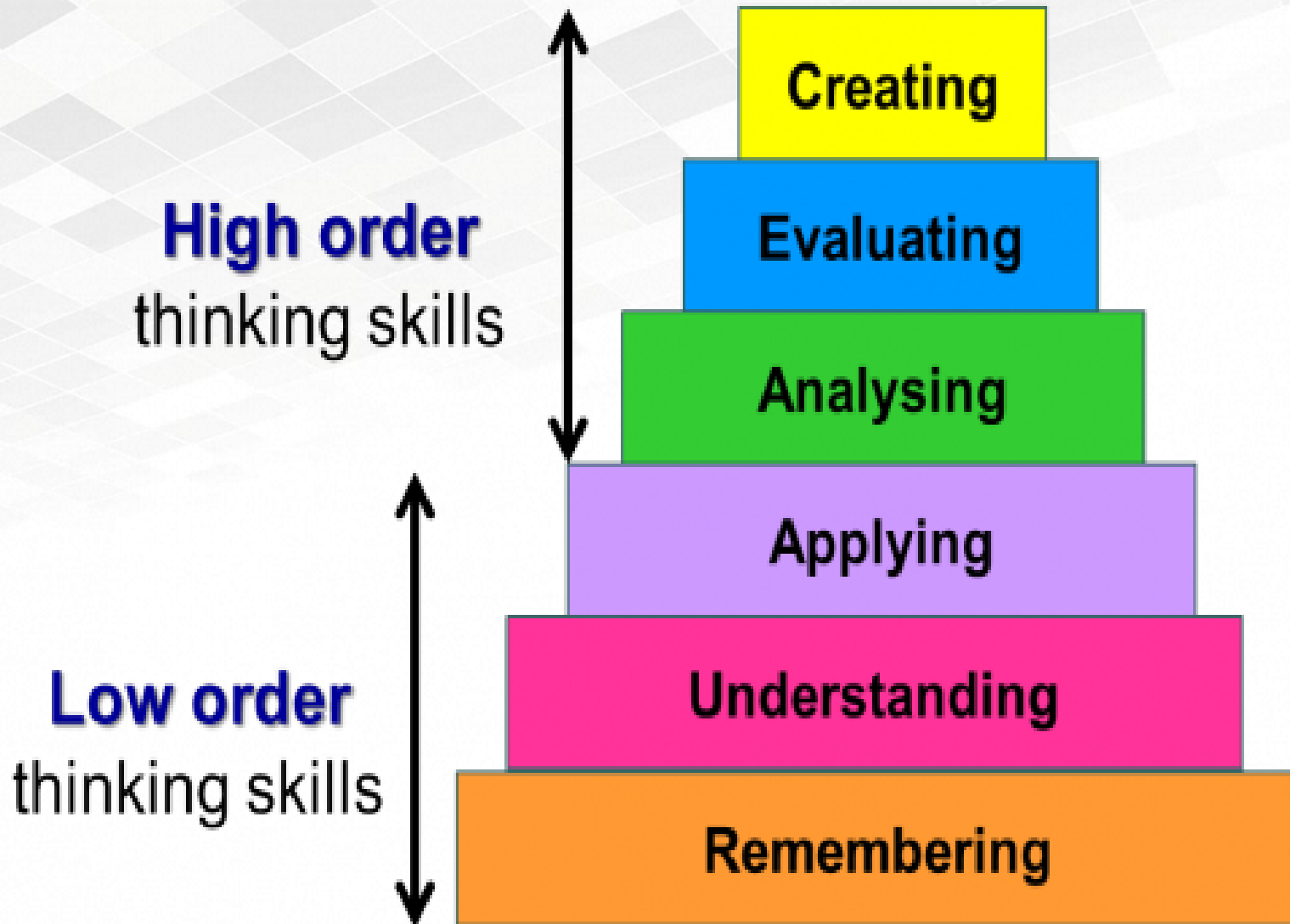| Assessing Products	Essays with prompts, criteria Projects, with criteria Portfolios, with criteria Demonstrations, exhibitions Investigations, Performances Paintings, Drama, Stories, with criteria Attitude Inventories, Survey Written tests Electronic databases, stories Portfolios and Hyper card stacks

1956

- Evaluation
- Synthesis
- Analysis
- Application
- Comprehension
- Knowledge

2001

- Create
- Evaluate
- Analyze
- Apply
- Understand
- Remember

Noun ⟶ to Verb Form

Anderson and Krathwohl's Revision of Bloom's Taxonomy

The Knowledge Dimension	The Cognitive Process Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge						
Metacognitive Knowledge						

Categories of Assessment Procedures

Type 1 . **By Method of Development**
            Teacher-constructed
            Standardized
Type 2. By Level of Formality
            Informal
            Formal
Type 3. By Instructional Purpose
            Formative (Diagnostic)
            Summative

Categories of Assessment Procedures

Type 4. **By Type of Grading Standard**

Criterion-Referenced

Norm-Referenced

Type 5. By Item Formal

Objective

Constructed-Response

Type 6. **By Degree of Authenticity**

Performance

Traditional

General Principles of Assessment

1. **Clearly specifying what is to be assessed has priority in the assessment process**

2. An **assessment procedure should be selected because of its relevance to the characteristics or performance to be measured.**

3. **Comprehensive assessment requires a variety of procedures.**

(a) **multiple choice and short answer tests** of achievement are useful for measuring knowledge, understanding, and application outcomes

(b) **essay tests and written projects** are needed to assess the ability to organize and express ideas

(c) **projects** that require students to formulate problems , accumulate information through library research, or collect data (e.g. through experimental observations or interviews) are needed to measure certain skills in formulating and solving problems.

(d) **Observational techniques** are needed to assess performance skills and various aspects of student behavior.

(e) **self-report techniques** are useful for assessing interests and attitudes.

4. Proper use of assessment procedures requires an awareness of their limitations

- Sampling error

- Chance factors influencing assessment results, such as guessing on objective tests, subjective scoring on essay tests, errors in judgment on observation devices, and inconsistent responding on self-report instruments (e.g, attitude scales).

- Incorrect interpretation of measurement results

5. Assessment is a means to an end, not an end in itself.

- The use of assessment procedures implies that some useful purpose is being served and that the user is clearly aware of this purpose.

- To blindly gather data about students and then file the information away is a waste of both time and effort.

- Assessment is best viewed as a process of obtaining information on which to base educational decisions.

Understanding Validity and Reliability in Classroom, School-Wide Assessments for Teachers and Principal

Validity

- "**Validity** refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests...

- It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself" (Messick, 1989, p. 9)

- Many sources define validity as the **test measuring what it claims to measure, but we are actually talking about the ability to make inferences from the data and that these inferences are indeed correct.**

Validity

- The most important factor in **test development** is to be sure you have created an assessment that allows you to make appropriate **inferences** regarding a **child's performance** in a **content area**, and now to feel confident that you can **attribute a teacher's contribution to this performance**.

- This requires you to begin with "**a clear statement of the proposed interpretations and uses**" (Kane, 2006, p.23) of the assessment.

- If we are testing a child's knowledge of 5th -grade social studies concepts, but the child's reading level is at the 1st -grade level, then the resulting score would be more a reflection of the child's reading ability than the child's knowledge of social studies;

- therefore, we would not be able to make a valid inference from the test score and the test would not have validity for this child.

- It would then follow that we cannot make a valid inference about the teacher's contribution to a child's growth in the social studies content if the child's reading level is convoluting the final score.

Three basic types of validity

- construct validity
- criterion validity
- content validity
- **Construct validity** "refers to the skills, attitudes, or characteristics of individuals that are **not directly observable** but are **inferred** on the basis of their observable effects on behavior" (Martella, Nelson, and Marchand-Martella, 1999, p. 74).

Construct validity

- The extent to which a measurement device can be shown to measure a **hypothetical construct**. Examples of construct are intelligence, creativity, reading comprehension capabilities, or mathematical competence.

- We can find many differing views about what skills constitute creativity, but most would be in agreement about what skills demonstrate mastery of two-digit addition.

Criterion validity

- Criterion validity is "used when test scores can be related to a criterion.

- The criterion is some **behavior** that the test scores are **used to predict**" (Allen & Yen, 2002).

- Criterion validity can also be called **concurrent validity**, where a relationship is found between two measures at the same time.

- Since this is seldom used in today's testing environment, we will only focus on criterion validity as it deals with the **predictability of the scores**.

Criterion validity

- The SAT (standardized test) is a good example of a test with predictive validity when the test scores are highly correlated with success in college, a future performance.

- To determine both construct validity or criterion validity, calculations and examination of correlations or other statistics are used.

- On the other hand, to determine content validity no statistical calculations are used (Allen & Yen, 2002).

Content validity

- "refers to the **representativeness** of the sample of items included in measurement devices" (Martella et al., 1999).

- The degree to which a measurement actually reflects the variable it has been designed to measure.

- In other words, **does the test content appropriately represent what was taught**?

Content validity

- "The key ingredient in securing content-related evidence of validity is human judgment" (Popham, 2000, p. 96).

- The human judgment in any achievement testing would be the teachers who are teaching the content, and to be sure you have stronger validity evidence, you should include several teachers (and content experts when possible) in evaluating how well the test represents the content taught.

Reliability

- **Reliability** is "the desired consistency (or reproducibility) of test scores (Crocker & Algina, 2008).

- We would want to know if the child took the test again, the **score would be similar or consistent over multiple testings** (Brennan, 2006).

- In test design, reliability is referring to the confidence you have that the test score earned is a good representation of a child's actual knowledge of the content, or is a good representation of a child's true score if there were no such thing as measurement error and we could design a perfect test.

1. **Coefficient of stability (test-retest)**

- the coefficient of stability (test-retest) is a measurement of how the **scores from one test remain fairly similar** when a group of students takes the same test twice, not just for the individual student but also in how they perform in relationship to other scores.

- In the more reliable measures, the student's individual scores remain fairly consistent, as well as the rank order of how well students perform on one test will remain close to the same on the second measure.

2. Coefficient of equivalence (alternate form)

- is a measurement of how the scores on one test remain similar to the scores on a second, or parallel, test given around the same time.

- A measure of the magnitude of relationship between participants' scores on two comparable forms of the measurement device.

- The coefficient of equivalence is also dependent on the rank order of how well students perform, having more reliability when the rank order of scores on one measure remains close to the same rank order on the second measure.

3. Coefficient of internal consistency

- is a measure of how well the items are working together to measure the same concept.

- A measure indicating the magnitude of relationship between participants' scores on a single administration of the measurement device usually assessed by comparing two parts (e.g. odd and even items, first half to second half).

- often computed by what is called "split-half correlation" where there are multiple ways for a test to be divided into two separate scores thus yielding the coefficient of equivalence (Martella et al., 1999; McMillan & Schumacher, 1997).

Southeast Asian Ministers of Education Organization
Regional Centre for Education in Science and Mathematics

Error of measurement

- Error of measurement or Random error

- Fluctuations in scores because the measurement device does not measure an attribute the same way every time; unknown and unrepeatable causes of variability in task performance over time and context

Error of measurement

- Typically, measurement error is reported as a plus/minus number that is then added to your observed score to give the range of possible values of the student's true score.

- For example, if a student scores a 75% with a measurement error of +/-3, this tells you that 68% (one standard deviation from the mean in a normal distribution) of the time the child's true score would range between 72% and 78%, and 95% (2 standard deviations from the mean in a normal distribution) of the time the child's true score would range between 69% and 81%

- A teacher-developed test can be made more **reliable and have less measurement error** by making sure you have written clear directions so students know exactly what is expected of them from the assessment;

- that you have appropriate questions that clearly measure the content taught and are not confusing students with the wording or biased to any subgroup of students; and

- that you have gathered appropriate feedback from colleagues and students who have read through your assessment or taken the assessment to provide feedback to its clarity (Brennan, 2006; Nhouyvanisvong, 2015; Popham, 2014).

Interrater Agreement

- The degree to which two or more evaluators using the same rating scale give the same rating to an identical observable situation.

- With rubrics or performance assessments, you will typically be concerned with the internal consistency expressed as a coefficient, most often called interrater reliability.

Interrater Agreement

- Interrater reliability – the consistency between evaluators in the ordering or relative standing of performance ratings.

- This form of reliability is to make sure observers are measuring the same variables and are consistent in how they rate each of the variables (McMillan & 9 Schumacher, 1997; Stemler, 2004).

Steps in Designing a Local Assessment

1. Identify the test purpose or test objective.

2. Construct a Table of Specifications/determine item format.

3. Construct initial pool of items.

4. Review test items with colleagues and students; then revise as necessary.

5. Pilot test your assessment; then revise as necessary.

6. Administer and score your assessment; review and revise as necessary.

References

Assessment Reform Group website: http://www.assessment-reform-group.org/

Black, P. (1998). Formative assessment: raising standards inside the classroom. *School Science Review, 80*(291), 39-46.

Black, P. & Wiliam, D. (1998). *Inside the Black Box* (online version available at http://blog.discoveryeducation.com/assessment/files/2009/02/blackbox_article.pdf).

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning.* Thousand Oaks, CA: Corwin Press.

Glasson, T. (2009). *Improving student achievement: A practical guide to assessment for learning.* Carlton, Australia: Curriculum Council.

Groves, R. (2009). *Talking about teaching: Teachers' learning in a professional development context.* Saarbrucken, Germany: VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG.

Hattie, J. (2008). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement.* London: Routledge.

Heppell, S. (2008).Changing possibilities for 21st century teaching and learning. Presentation at Curriculum Council of Australia Annual Conference, November 2008. Downloaded in October 2009 from http://www.curriculum.edu.au/conference/2008

- https://cms.azed.gov/home/GetDocumentFile?id=57f6d9b3aadebf0a04b2691a