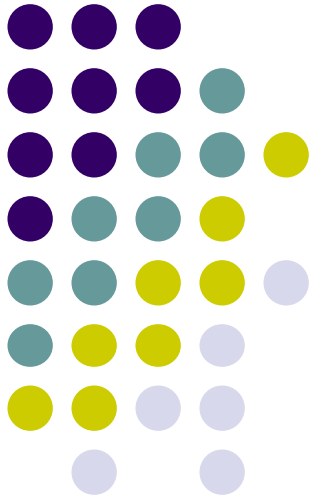
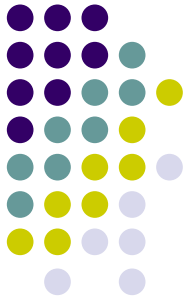
Improving Assessment Plans

Finding Solutions to Common Methodological Issues

Dr. Steve Atkins
California University



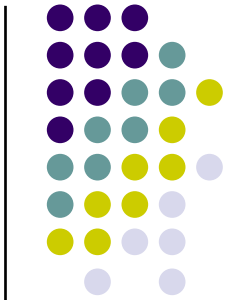
Review: Writing the Outcome



Student Learning Outcomes state what students are expected to *know* or *be able to do* upon completion of a program/utilizing an administrative service.

- SLO statements should focus on the expected **abilities**, **knowledge, values, and attitudes** of a student after they have completed your program.
- SLO statements should focus on the results of learning and not the learning processes.
- SLO statements should be stated such that the outcome can be measured by more than one assessment method.

Review: Writing the Outcome

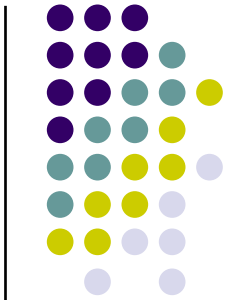


Student Learning Outcomes…

- ✓ Are specific, observable behaviors evidenced by students who have achieved your educational objectives.
- ✓ Learning outcomes are stated operationally, and describe the observable evidence of a student's knowledge, skill, ability, attitude or disposition.
- ✓ State clearly each outcome you are seeking: How would you recognize it? What does it look like? What will the student be able to do?

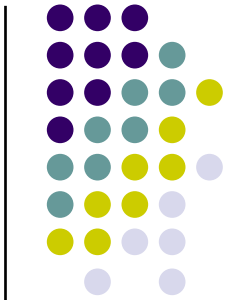
https://academicprograms.calpoly.edu/content/assessment/assessplanguide#define

Review: Writing the Outcome



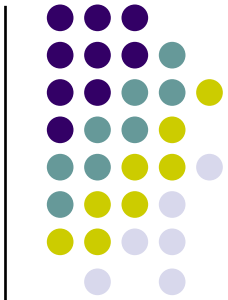
You should use **concrete verbs** like define, classify, operate, formulate, rather than passive verbs like “be exposed to” or vague verbs like understand, know.

Sample SLO Example:



- Given a set of data, the student will be able to compute the standard deviation.
  - ✓ Condition – given a set of data
  - ✓ Behavior – the student will be able to compute the standard deviation
  - ✓ Criterion – (implied) – the number computed will be correct

#1 Provide greater specificity in Student Learning Outcomes (SLOs)

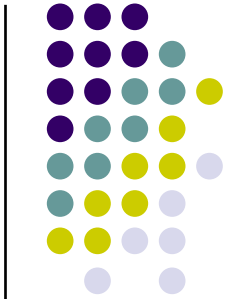


SLO statements should be **distinctive** and **not generic**.

Example of a generic outcome:

"Students completing an AS in Architectural Design will be practiced in design skills."

#1 Statements should be neither too specific, nor too general



For example, if a university's Education Department were writing statements…

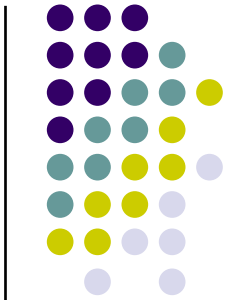
- **TOO SPECIFIC**: "Students will be able to compose true/false statements."
- **TOO GENERAL**: "Students will be able to prepare a good test."
- **MORE APPROPRIATE**: "Students will be able to apply alternative forms of evaluation in classroom teaching situations that yield valid results."

EXAMPLE

TOO GENERAL: "Students completing the undergraduate program in Hypothetical Engineering will have knowledge of engineering principles."

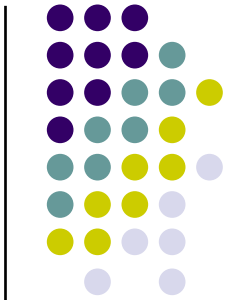
This is a weak statement because it does not specify which engineering principles a graduate from the program should know (vague). Also, it does not define what is meant by "have knowledge." Are they supposed to be able to simply **define** the principles, or be able to **apply** the principles, etc.?

BETTER: "Graduates will be able to apply the principles of engineering design, formulate requirements and constraints, following an open-ended decision process involving tradeoffs."



This is a much better learning outcome statement for two reasons. First, the specific requirements are listed and second, the level of competency is also stated. A student must be able to apply and to demonstrate the listed engineering principles.

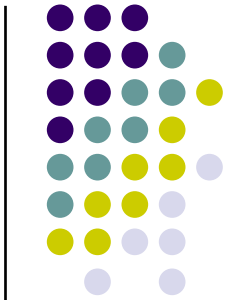
BETTER: "Graduates will **be able to apply** the principles of engineering design, **formulate** requirements and constraints, following an open-ended decision process involving tradeoffs."



This is a much better learning outcome statement for two reasons:

1. The **specific requirements** are listed
2. The **level of competency** is also stated. A student must be able to **apply and to demonstrate** the listed engineering principles.

EXAMPLE



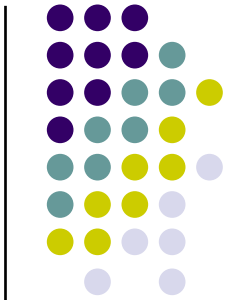
TOO GENERAL: “Students should be able to independently design and carry out research.”

The problem with this is that the statement **does not specify** the type or quality of research to be done.

BETTER: "Students should be able to independently design and carry out experimental and correlational research that yields valid results."

Here the **standard for students to aim for is clear and specific** enough to help faculty agree about what students are expected to do. Therefore, they **should be able to agree reasonably well about whether students have or have not achieved the objective**. Even introductory students can understand the sentence, even if they don't know exactly what experimental and correlational research methods are.

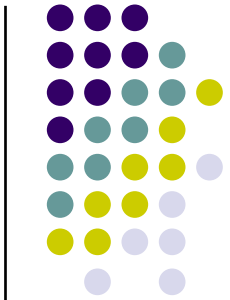
#2 Alignment Between an Outcome and its Measures



- Make sure measures are **precise** and described with **sufficient specificity** (e.g. “students will write a report” is imprecise.)

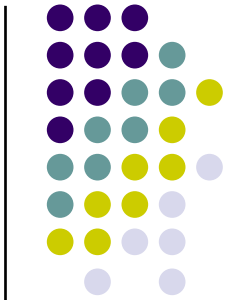
- Provide a **clear connection** to program outcome.

#3 Using Grades as Outcomes Measures



- Course grades provide a student with general feedback about his or her overall performance in a course.
- They often consist of exams, projects, and other direct measures. However, they also **frequently include variables** such as neatness, attendance, tardiness, and extra credit, **which do *not* reflect what a student has learned and are therefore irrelevant to SLO measurement.**

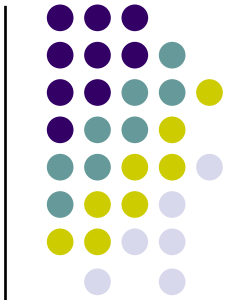
#3 Using Grades as Outcomes Measures



Innapropriate use of grades as outcomes measures:

- “80% of the class will have a minimum grade of C on the exit exam”
- “70% of Students will receive a grade of C or higher on a comprehensive exam”

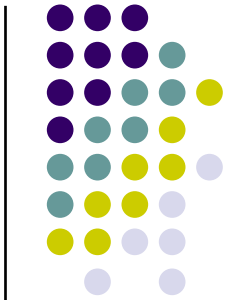
#3 Using Grades as Outcomes Measures



BETTER:

"Students enrolled in the 'Capstone in Economics' course will be required to complete an essay assignment related to Microeconomics. A faculty committee will assess the performance of a random sample of students, using the attached rubric. A minimum of 70% of students completing the essay will achieve ratings of satisfactory or higher on the rubric."

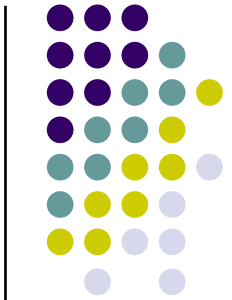
#3 Using Grades as Outcomes Measures



BETTER:

“Based on an employer survey, which is administered once every two years, at least 80% of the employers will be satisfied with the knowledge of ethics and conduct of our student interns.”

Grades Not Actionable



Grade Report

- George 91
- Steve 100
- Jean 92
- Mark 84
- . . . .

Average 78

(N = 24)

Adapted from: May, G. (2008). *Using Assessment Data to Improve Teaching and Learning.* Clayton State University: School of Business.

Performance Outcome: Actionable Data

Critical Thinking Analysis

N = 30

Performance Element	Scoring Satisfactory or Better	%
Purpose	29	97%
Information	19	63%
Assumptions	25	83%
Implications	27	90%
Points of View	26	87%
Conclusion	16	53%

May, G. (2008). *Using Assessment Data to Improve Teaching and Learning.* Clayton State University: School of Business.

Don't Have Time?

Communication Strategy Analysis

Element	Unsatisfactory	Satisfactory	Good
Purpose			
Information			
Assumptions			
Implications			
Points of View			
Conclusions			

May, G. (2008). *Using Assessment Data to Improve Teaching and Learning.* Clayton State University: School of Business

Scoring Rubric For Tennis Serve			
	Excellent (2 points)	**Acceptable** (1 point)	**Poor** (0 points)
Accuracy	Hits target area at least 80% of the time	Hits target area about 50%-80% of the time	Hits target area less than 50% of the time
Effort	Puts forth good effort and tries to serve correctly	Shows moderate interest in learning how to serve	Shows little or no interest in learning the correct way of serving

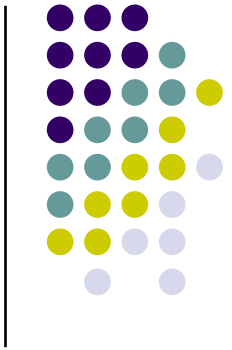
Adapted from: Saret, L. (200 8*). Learning Outcomes*. Oakton Community College

Class Participation Rubric

	3	2	1
Participates in class discussions	Almost Always	Occasionally	Almost Never
Asks relevant questions in class	Almost Always	Occasionally	Almost Never
Participates in online discussion	Almost Always	Occasionally	Almost Never
Offers questions or comments via e-mail	Almost Always	Occasionally	Almost Never
Attends class	Almost Always	Occasionally	Almost Never
Arrives on time/stays for entire class	Almost Always	Occasionally	Almost Never

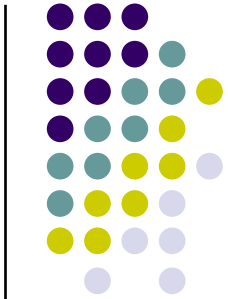
Adapted from: Saret, L. (200 8*). Learning Outcomes*. Oakton Community College

Performance Rubric: Business/Management



Performance Characteristics			
Team's Customer Satisfaction Skills	Does not meet Expectations	Meets Expectations	Exceeds Expectations
Punctuality	Some team members missed appointments or did not return phone calls.	All team members arrived on time for appointments and returned all phone calls promptly.	All team members were always early.
Courtesy	Some team members were not respectful of firm employees	All team members were always courteous and respectful of all firm employees.	All employees felt that the team members were very courteous and respectful and fully elicited their ideas.
Communication	Some team members did not communicate clearly during meetings and phone calls.	The team members always communicated clearly during meetings and phone calls.	The team members always made an extra effort to make sure that they understood us and that we understood them.

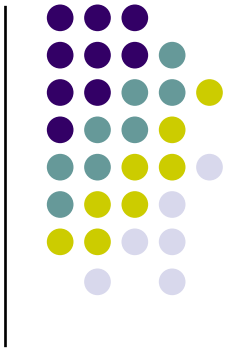
Rubrics



Advantages of using rubrics:

- Allow assessment to be more outcomes-focused and consistent
- Focus the teacher to clarify his/her criteria in specific terms
- Clearly show the student how their work will be evaluated and what is expected
- Promote fair and consistent grading
- Promote student awareness of about the criteria to use in assessing peer performance
- Provide useful feedback regarding the effectiveness of the instruction and help identify areas for improvement
- Provide benchmarks against which to measure and document progress

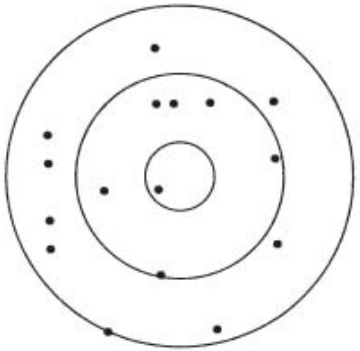
Rubrics



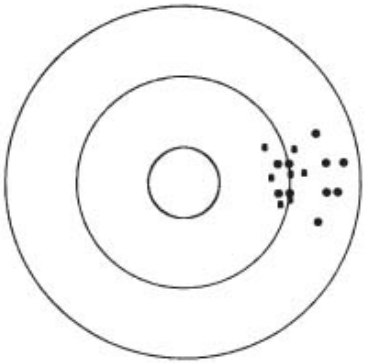
- State if **multiple reviewers** will be used (e.g. two reviewers will score the rubrics)
- Include a brief description about **inter-rater reliability**
- State that reviewers will be given some kind of **training** (calibration of the rubric)

#4 Validity and Reliability

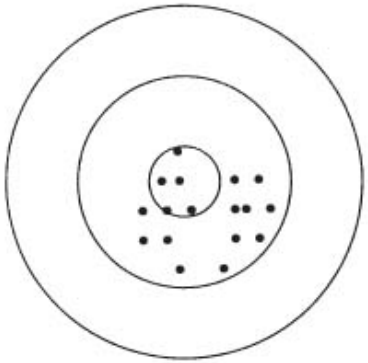
Think in terms of '*the purpose of tests*' and the '*consistency*' with which the purpose is fulfilled/met



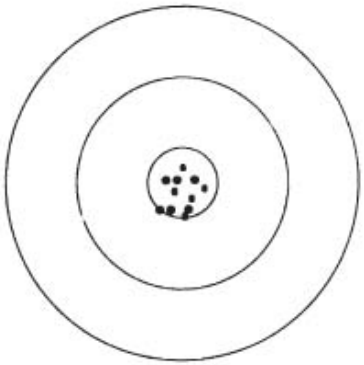
Neither Valid nor Reliable



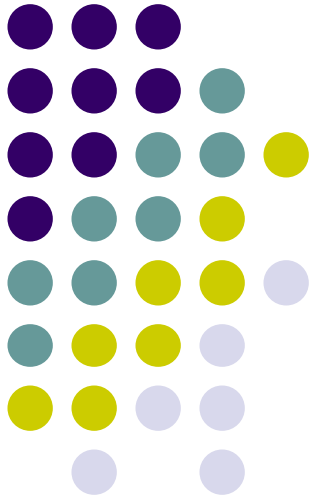
Reliable but not Valid



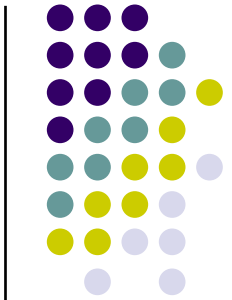
Fairly Valid but not very Reliable



Valid & Reliable

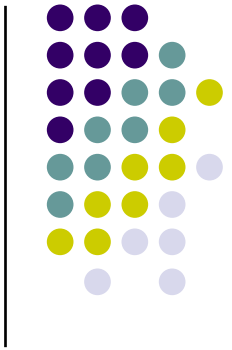


Validity



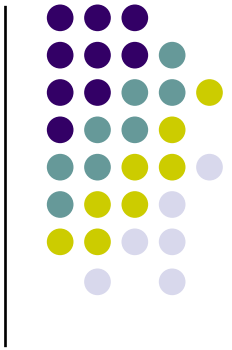
- Depends on the **PURPOSE** of the assessment tool
  - e.g. a ruler may be a valid measuring device for length, but isn't very valid for measuring volume
- Measuring what 'it' is supposed to
- Matter of degree (how valid?)
- Specific to a particular purpose!
- Must be inferred from evidence; cannot be directly measured
- **Begins with Learning Outcome statement**

Content Validity



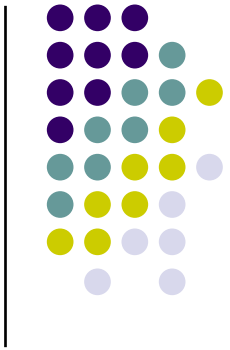
- How well elements of the test relate to the **content domain**?
- How closely content of **questions** in the test relates to content of the **curriculum**?
- Directly relates to outcomes!
  - ✓ Use a rubric
  - ✓ Table of Specifications
  - ✓ Validate assessment using experts in the field

Factors that can lower validity



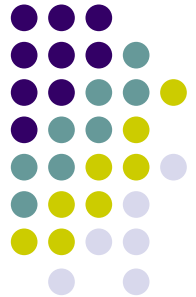
- Unclear directions
- Difficulty reading vocabulary and sentence structure
- Ambiguity in statements
- Inadequate time limits
- Inappropriate level of difficulty
- Poorly constructed test items
- Test items inappropriate for the outcomes being measured

Factors that can lower validity



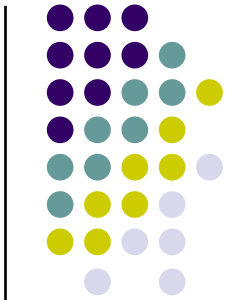
- Tests that are too short
- Improper arrangement of items (complex to easy?)
- Identifiable patterns of answers
- Teaching
- Administration and scoring
- Students
- Nature of criterion

Test Reliability



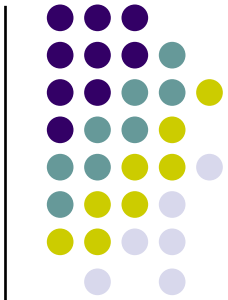
- **Measure of consistency** of test results from one administration of the test to the next
- When someone says you are a ‘reliable’ person, what do they really mean?
- Deals with the question of “how well can we depend on the score as a measure of student’s “real” learning?”
- A component of validity
- Length of assessment
- ***A reliable test produces similar scores across various conditions and situations, including different evaluators and testing environments***

Reliability



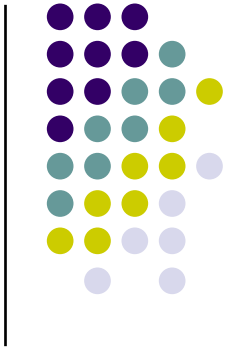
- **Inter-rater reliability** is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter-rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed.
- Use a rubric, at least 2 graders, and train them

How to improve reliability



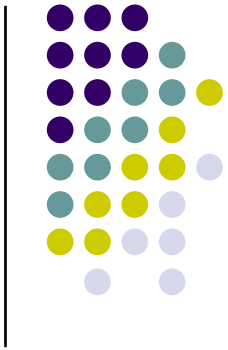
- Quality of items; concise statements, homogenous words (some sort of uniformity)
- Adequate sampling of content domain; comprehensiveness of items
- Longer assessment – less distorted by chance factors
- Developing a scoring plan (rubrics)
- Ensure **VALIDITY**

Validity and Reliability Checklist



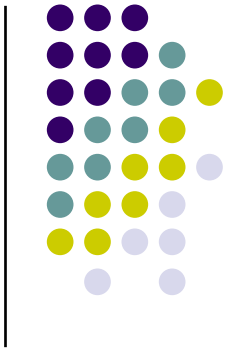
- ✓ Does the assessment adequately evaluate academic performance relative to the desired outcome? (Validity)
- ✓ Does the assessment method adequately address the knowledge, skills, abilities, and values associate with the intended outcome? (Domain validity)
- ✓ Will the data accurately represent what the student can do in an authentic or real life situation? (Authentic assessment and validity)

Validity and Reliability Checklist



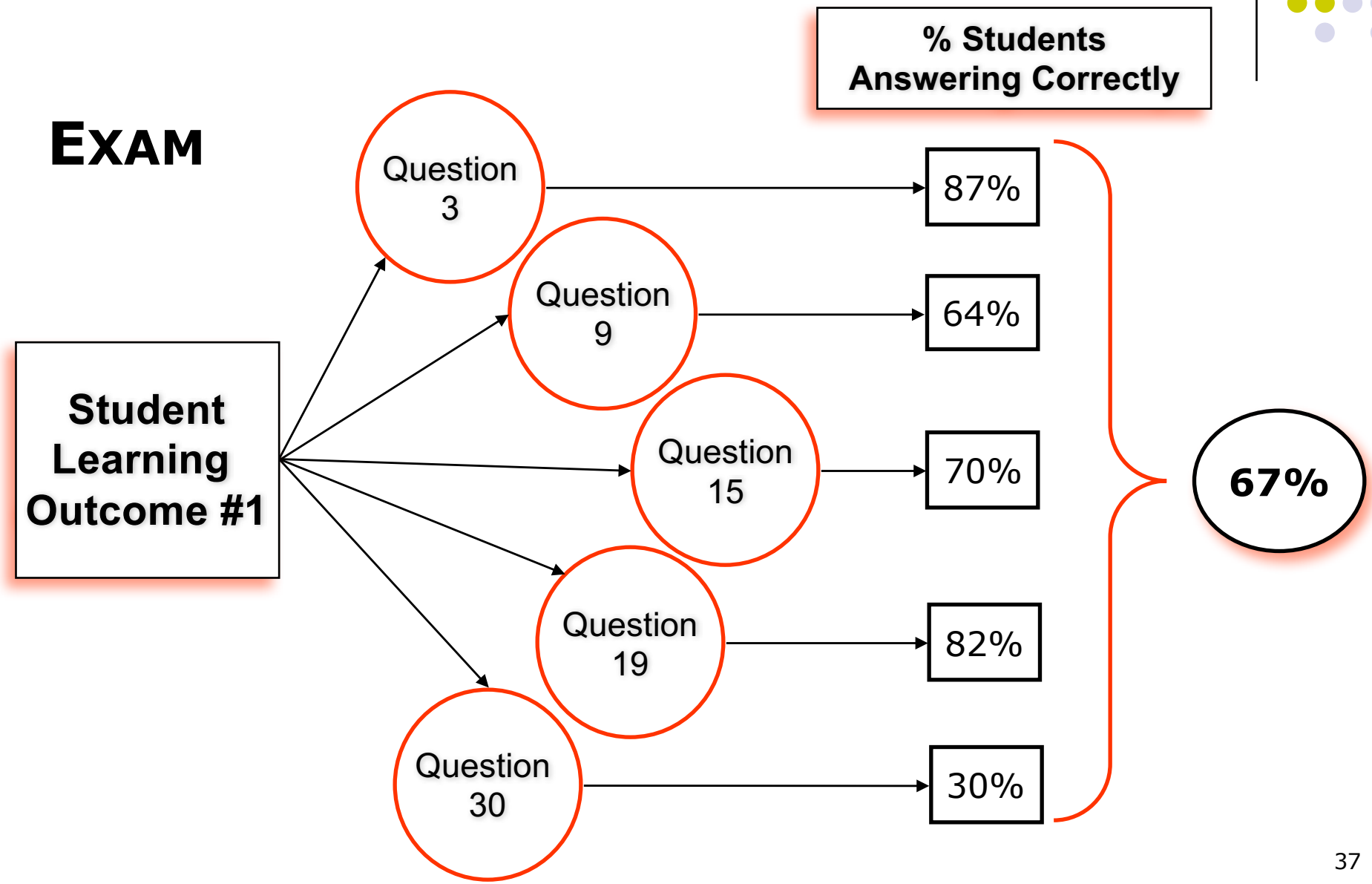
- ✓ Is the grading scheme consistent; would a student receive the same grade for the same work on multiple evaluations? (Reliability)
- ✓ Can multiple people use the scoring mechanism and come up with the same general score/ (Reliability)
- ✓ Does the assessment provide data that is specific enough for the desired outcomes (alignment with SLO)
- ✓ Is the assessment summative or formative – if formative does it generate diagnostic feedback to improve learning? If summative, is the final evaluation built upon multiple sources of data/ (AAHE good practice)

Validity and Reliability Checklist



- ✓ Are the intended uses for the assessment clear? (grading, program review, both)?
- ✓ Have other faculty provided feedback?
- ✓ Has the assessment been pilot-tested?
- ✓ Will you provide the students with a copy of the rubric or assignment grading criteria?
- ✓ Will you provide the students example of model work?

#5 Course Embedded Measures



May, G. (2008). *Using Assessment Data to Improve Teaching and Learning.* Clayton State University: School of Business

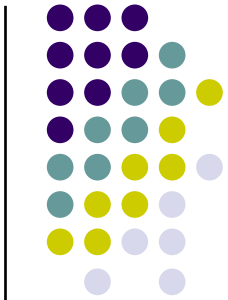
Embedded Outcomes: Actionable Data

Students will be able to know, comprehend, and apply the following four principles of effective communication:

Student Learning Outcomes	Number of Exam Questions	% Students with Correct Answers
1. Building Goodwill	5	67%
2. Adapting Message to Audience	7	91%
3. Making Writing Easier to Read	8	58%
4. Using a Process to Plan Compose and Revise	5	84%

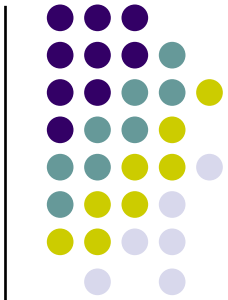
May, G. (2008). *Using Assessment Data to Improve Teaching and Learning.* Clayton State University: School of Business

#6 Capstone Course/Project and Portfolios



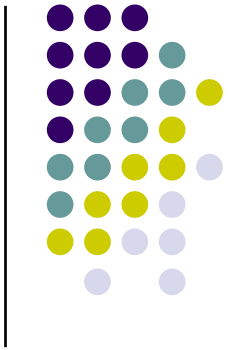
- Assessment of capstone projects and Portfolios should have the following elements:
  - Multiple readers/scorers (A minimum of two faculty members)
  - Clear and concise rubrics
  - A statement regarding a method for measuring inter-rater reliability for consistency of measurement.

#7 Writing Samples or Oral Presentations



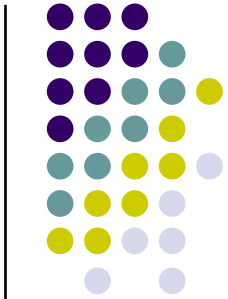
- Assessment of Writing Samples and Oral Presentations should have the following elements:
  - Multiple readers/scorers (A minimum of two faculty members)
  - Clear and concise rubrics
  - A statement regarding a method for measuring inter-rater reliability for consistency of measurement.

#8 Common Test Item Bank



- Such items should have the following characteristics:
  - Some effort to judge reliability and validity, or selection of items from a test bank with already proven reliability and validity
  - Item analysis so that appropriate item discrimination may be ascertained

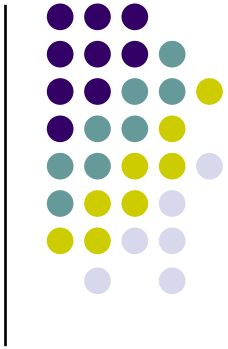
#9 Indirect Measures



Surveys are the most prevalent indirect measures appearing in Cal U assessment plans.

- Departments need training in survey design and minimally, need to be reviewed by assessment personnel prior to their use.
- There is a marked difference between a student's report that s/he has become a better writer and a skill test to measure quality of writing. Thus, student perceptions of their abilities via survey research should not be used as the sole or predominant means of measuring learning outcomes.

Direct and Indirect Assessment of Learning

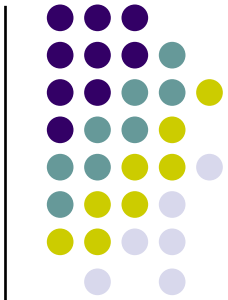


- *Direct* evidence of student learning is tangible, visible, and compelling evidence of exactly what students have and have not learned.

- *Indirect* evidence consists of proxy signs that students are probably learning. Indirect evidence is less clear and convincing.

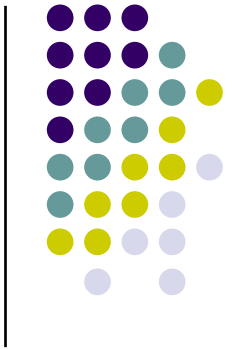
Adapted from: Sweet, J. and Meents-Decaigny. (2015) *Direct assess of student Learning* .DePaul University

Examples of Evidence: Direct or Indirect?



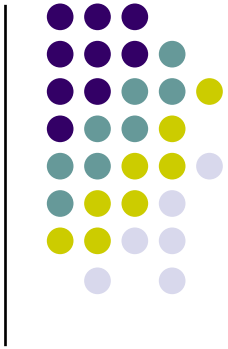
For each of the following examples, indicate whether you think this is direct or indirect measurement:

Examples of Evidence: Direct or Indirect?



1. Student reflections evaluated by your Assessment Committee
2. Student exams, papers, projects, computer programs
3. Observations of student interactions with a client during internship
4. Student musical performance
5. Survey that asks students to report their learning for particular concepts
6. Asking students or alumni how well they thought they learned

Examples of Evidence: Direct or Indirect?



7. Tracking graduate school or job placement rates
8. Retention and graduation rates for students in a particular program
9. Teaching evaluations
10. Site Supervisor ratings of student performance during an internship
11. Survey that asks students to respond to a scenario using content learned in class
12. Course grades