# Planning and Assessing Student Learning Outcomes for Statistics

Carl Lee
Central Michigan University
Mt. Pleasant, MI 48849
Carl.lee@cmich.edu
Phone: (989) 774-3555



# Outline

- Teaching Vs. Learning, Teachers Vs. Students
- A Framework
- Designing an Assessment Plan for Statistics Major
- Some Tools for Classroom Assessment
- Course Level Assessment – Performance-based assessment
- Course Level Assessment – Intro to Statistics
- Assessing Misconceptions
- Assessing Statistical Literacy, Reasoning and Thinking



# Activity
# Students' Wish List



# Activity
# Your Wish List



# Activity
## Teaching Goals Inventory (Angelo, Cross (1993)

Think of one of your courses currently you are teaching:_____

Think of the program or department where you belong to: _____

**Rate each of the following six major goals using** 5: Essential, 4: Very Important, 3: Important, 2: Some what important, 1: Little Important, 0: Not Applicable

| Teaching/Learning Goals | Rate for a course | Rate for a program |
|---|---|---|
| (a) Higher-Order Thinking Skills | _____ | _____ |
| (b) Basic Academic Success Skills | _____ | _____ |
| (c) Discipline-Specific Knowledge & Skills | _____ | _____ |



# FIND SOME COMMON GROUNDS

We wish students to learn high order thinking & problem solving skills, core academic knowledge, life long learners.

Students wish to prepare for career, care more about personal development, mostly only willing to work on grade-related activities that we design for them.

**WHAT SHOULD WE DO?**

## Do students' learning goals conflict against teacher's teaching goals?

**Not really!**

Most college students may not have enough experience to realize what else will be needed for their career/future other than the core knowledge.

If critical thinking, problem solving, academic core knowledge, life-long learning really matter, then let's make sure students are properly informed, taught, and assessed.

This is what student learning assessment all



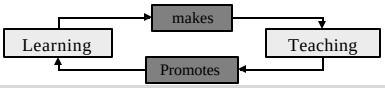
## Nine Key Traits for Good Teaching and Learning on which Students and Instructors Agree

1. Knowledge of the Subject/Discipline
2. Course Preparation and Organization
3. Clarity and Understandability
4. Enthusiasm for Subject/Teaching
5. Sensitivity to and Concern for Students' Level and Learning Progress
6. Availability and Helpfulness
7. Quality of Examinations
8. Impartiality in Evaluating Students' Work
9. Overall Fairness to Students

(Ewell, P.T., and D.P. Jones, 1996)



## *Teaching Vs. Learning*

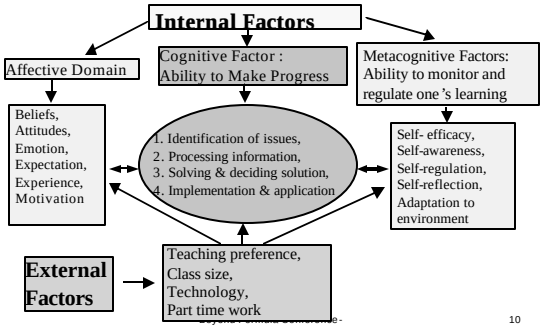



- The way we teach mostly likely reflects the way we learn.
- The 1995 APA Guideline (1995): Learning is a process of creating meaningful representations of knowledge through internally mediated processes including self-awareness, self-questioning, self-monitoring, and self-regulation.
- The way we learn, generally speaking, is related to three major dimensions:

1. The cognitive dimension
2. The affective dimension
3. The metacognitive dimension

## *Four Dimensions of Learning Process*

(Leahey & Harris (1997; APA Guideline (1995); Sternberg (1999); Hartman (2001))






## Cognition Vs. Metacognition

(Hartman 2001)

Cognitive strategies enable one to MAKE progress – to build knowledge.

Metacognitive strategies enable one to MONITOR and IMPROVE one's progress – to evaluate the understanding and make plans for improvement.

Studies have shown that cognitive ability has little relation with the acquisition of metacognition. (e.g., Alexander, 1995; Schraw, 2001).



## Expert Learners Vs. Novice Learners

Literatures have shown that (Larken, et al (1983), Margolis (1987), A'Andrade (1995))

- The way experts learn is very different from novices.
- Experts have sufficient domain knowledge. Solving problems by recognizing patterns or similar situations they have experienced in their mental models is often sufficient and successful for solving problems.
- Novices have little domain-specific knowledge. Hence, they tend to use more general problem-solving strategies that are unrelated to domain knowledge (such as memorizing, breaking down to smaller problems, procedural strategies, connecting to daily experiences), until they have built sufficient subject matter knowledge.
- Familiar context enables the access of appropriate mental model for solving problems.

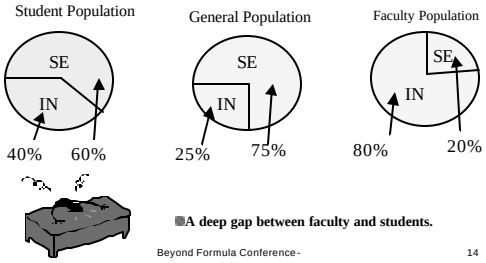
## Today's College Students

- Change Magazine (1993)- C. Schroeder: New Students – New Learning Styles
  - Study 4000 students using Myers-Briggs Type Indicator (MBTI)
  - **Learning Preference** --- Sensing Vs. Intuition (S Vs. N) and Extroversion Vs. Introversion (E Vs. I):
  - **Preference of Sensing or Extraversion (SE):** direct, concrete experiences, moderate to high degrees of structure, linear learning, practical, uncomfortable with abstract ideas, complex concepts, low tolerance of ambiguity, less independent in thought -- Practice-to-theory learning route
  - **Preference of Intuition or Introversion(IN)** : concepts,



## What is your learning style?

According to Schroeder (1993):




**A deep gap between faculty and students.**



## What is Student Learning Assessment

The activity of gathering and interpreting information related to students' achievement of learning goals/objectives at various stages of students' learning processes throughout the educational period and at graduation, as well as their performance in the workforce after graduation.

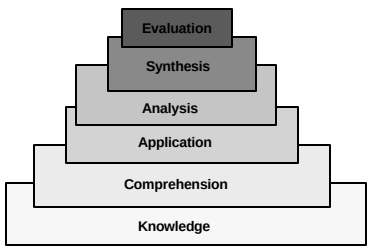


## What is Student Learning Assessment Four domains

- ***Knowledge domain***: the needed content knowledge in the area of study - **Cognition**
- ***Skills domain:*** the skills required for learning the content knowledge; For example, in a chemistry class, students should be able to use the lab equipment properly and safely – **Metacognition and Cognition**
- ***Attitudes and beliefs:*** this domain concerns the development of particular attitudes or change of beliefs about a discipline or a profession – **Affect**
- ***Values and ethics:*** Sound professional ethics and a commitment to life-long learning as well as good citizenship are important outcomes that a liberal arts



## Six Levels of Cognitive Taxonomy






## Important Trend in Assessment

| Emphasis of Assessment | Cognitive Views |
|---|---|
| View of learner | Active, constructing knowledge environment |
| Scope of assessment | Integrated and cross-disciplinary |
| Beliefs about knowing | Application and use of and being skilled facts and skills knowledge |
| Emphasis of instruction | Attention to metacognition, and assessment effective materials motivation, self-determination |

## Important Trend in Assessment

| Emphasis of Assessment | Cognitive Views |
|---|---|
| Characteristics of assessment | Authentic assessments on multiple-choice, contextualized problems that are short answer relevant and meaningful, emphasize higher-level thinking, may not have a single correct answer, have common standards known in advance, and are not speeded. |
| Frequency of assessment | Samples over time (portfolios) which provide basis for assessment. |
| Who is assessed | Assessment of group process skills on collaborative tasks which focus on distributions over averages. |
| What is assessed | Multidimensional assessment that learner recognizes the variety of human abilities and talents, malleability of student ability, and that IQ is not fixed |



## Characteristics of Good Assessment

- Good assessment information provides accurate estimates of student performance and enables teachers or other decision makers to make appropriate decisions.
- Good assessment information is its consistency, or reliability. Will the assessment results for this person or class be similar if they are gathered at some other time or under different circumstances or if they are scored by different raters?
- The content of the tests (the knowledge and skills assessed) should match the teacher's educational objectives and instructional emphases.
- The test items should represent the full range of knowledge and skills that are the primary targets of instruction.
- Expectations for student performance should be clear.
- The assessment should be free of extraneous factors which



## A framework for rethinking assessment

What the world needs:
Knowledge, Skills and Perceptions

Which are translated into Programs:
Major, Minor, Graduate, Certificates, Concentrations

Each has a list of
Learning Objectives and expected outcomes

That we define by topical areas and courses:



## A Framework

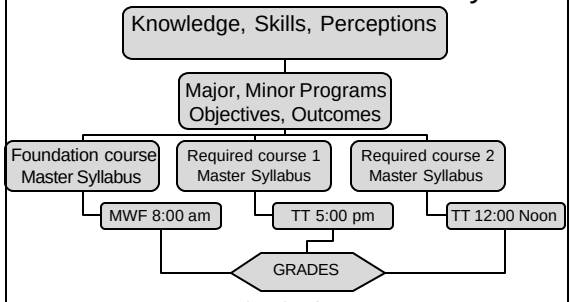
Each course has a master syllabus containing
Course description, prerequisites, course objectives, course policies, evaluation, schedule and bibliography

Which are implemented in
classrooms, labs, fields and so on

Where students learn
Knowledge, Skills and Perceptions

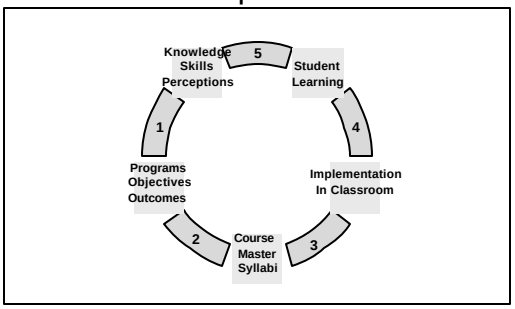


## A Typical Model of Educational Accountability





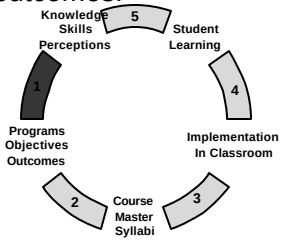

## The Model from Another Perspective



## Gap 1

What the world needs vs.
our perception of what's needed, as reflected in our programs, objectives, and outcomes:

How do we know our perceptions reflect reality?

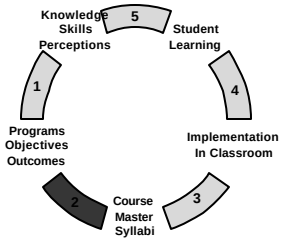





## Gap 2

our programs, objectives, and outcomes vs. our course master syllabi

How do we know our master syllabi accurately reflect our outcomes?

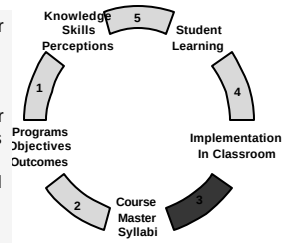





## Gap 3

our master course syllabi vs. classroom practice

- How do we know our implementation is true to our master syllabi?
- How do we know our classroom practice is consistent across multiple sections and instructors? instructors?

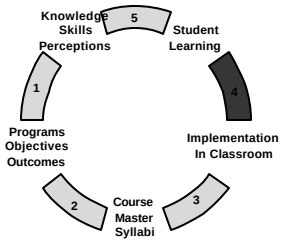





## Gap 4

What is taught vs. what students learn

How do we know that what students have "taken away" from their classroom experience is what we intended for them to learn?

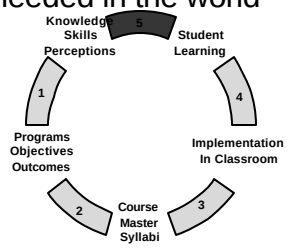





## Gap 5

what students have learned vs.
what is needed in the world

How do we know students are prepared to participate in their chosen careers with the "right" knowledge, skills, and perceptions?



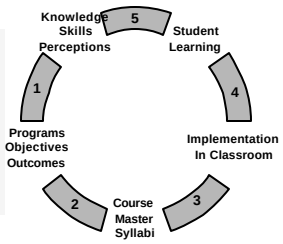



## Goal

To Close UP the Gaps

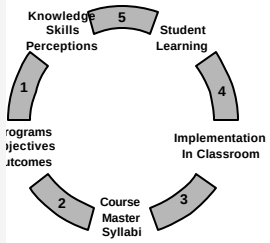
Focus:
assurance of student learning.

Key: articulate objectives and student learning outcomes.



# Goal
# To Close Up the Gap

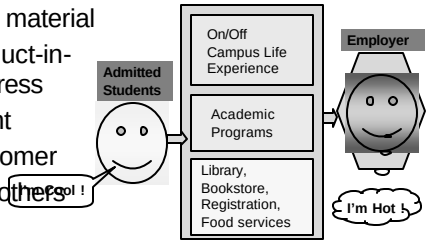
**Process: Process:**
1. define student learning outcomes
2. articulate content within each outcome
3. assign content to courses
4. develop delivery methods
5. measure learning against outcomes
6. adjust content or delivery to close the loop






# The Role for Students in Assessment

- Raw material
- Product-in-progress
- Client
- Customer
- And others






# Activity
# The Roles of Students

Describe the role(s) of students at an institution that you think students should be:

Raw material? Product-in-progress? Client?

Customer? And others (what)?



# The Role for Teachers in Assessment

- Ownership: Teacher owns. Bottom up, not top down
- Action: Continuing follow -up and support for change through coaching and mentoring
- Attitudes: Teaching Improvement, not teaching evaluation
- Others?



# The Role for Administration - Supporting Infrastructure

- Strong leadership support
- Staff development and training
- Create Environments that support experimentation and risk-taking
- Develop mechanism for rewarding – research resulted from assessment, teaching award, grant initiatives
- Other?



# Activity
# Design an Assessment Plan - Statistics

- Learning objectives
- Curriculum Alignment
- Evidence/Artifacts used to assess Student Learning Outcomes
  - Measurement tools and Expected Outcomes
- Dissemination of Information
  - Feedback loop
- Five year cycle

## 1. Objectives of Student Learning Outcomes for the program

**Domain: Knowledge**

Upon the completion of the program, students should be able to

K1
K2
K3
K4
K5

Each objective should not be too broad neither be too specific.

Example:

K1: interpret the quantitative information in daily life.



## Cognitive Action Verbs

- Identify, name, distinguish, define, describe, classify, demonstrate, translate, predict, generalize, explain, apply, analyze, recognize, judge, communicate, solve …..
- Focus on "What knowledge is essential to student success in your course or program" – Core curriculum.



## 1. Objectives of Student Learning Outcomes for the program

**Domain: Skills**

Upon the completion of the program, students should be able to:

S1
S2
S3
S4

**Domain: Perceptions, Values, Ethics**

Upon the completion of the program, students should

V1
V2
V3



## Affective Action Verbs

- Listening, responding, participating, seeking, demonstrating, relating, valuing, showing, realizing, …
- What dispositions are essential to student success in your course or program?



## 2. Curriculum Alignment of Student Learning Outcomes

- Where is the information introduced, emphasized, and/or reinforced in the courses required in the program?
- Use the format below to list the program Outcomes and the Core Courses. Fill in each cell with either:

  an **'I'** where the outcome is *introduced*,
  an '**E**' where the outcome is *emphasized*,
  and/or an '**R**' where the information is *reinforced*



| Learning Outcomes | Core Courses (I : Introduced R: Reinforced E: Emphasized) | | | | | | |
|---|---|---|---|---|---|---|---|
| K1 | | | | | | | |
| K2 | | | | | | | |
| K3 | | | | | | | |
| K4 | | | | | | | |
| K5 | | | | | | | |
| S1 | | | | | | | |
| S2 | | | | | | | |
| S3 | | | | | | | |
| S4 | | | | | | | |
| V1 | | | | | | | |
| V2 | | | | | | | |

## 3. Evidence/Artifacts used to assess Student Learning Outcomes over the 5 year period of this Plan

- What instruments will be used in each of the five years?
- When and where will they be administered in each of the five years?
- Which Student Learning Outcomes will be assessed during each of the 5 years?
- How will results be reported (e.g. percentages, ranks, state or national comparisons) for each of the 5 years?



| Outcomes to be Assessed | Instruments to be used | Expected Measures from the Instrument | Academic Year |
|---|---|---|---|
| | | | Year 1 |
| | | | Year 2 |
| | | | Year 3 |
| | | | Year 4 |
| | | | Year 5 |



## 4. Dissemination of Information over the 5 year period of this Plan

**When, where, and how**

will results be disseminated to stakeholders in each of the 5 years?

To instructors
To Students
To Department
To other related stakeholders



| Expected Measures from the Instruments (by Year) | Affected Stakeholders | Dates and locations for dissemination |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |

## Classroom Assessment Techniques (CATS) (Angelo & Cross, 1993)

- **What is classroom assessment?**

  Classroom assessment is both a teaching approach and a set of techniques. The approach assumes that the more you know about what and how students are learning, the better you can plan learning activities to structure your teaching. The techniques are mostly simple, non-graded, anonymous, in-class activities that give both you and your students useful feedback on the teaching-learning process.



## Classroom Assessment Techniques (CATS)

- **How is classroom assessment different?**

  Classroom assessment differs from tests and other forms of student assessment in that it is aimed at course improvement, rather than at assigning grades. The primary goal is to better understand your students' learning to improve your teaching.

## Classroom Assessment Techniques (CATS)

**How do I use Classroom Assessment Techniques?**

- Decide what you want to learn from a classroom assessment.
- Choose a Classroom Assessment Technique (CAT) that provides this feedback, is consistent with your teaching style, and can be easily implemented in your class.
- Explain the purpose of the activity to students, then conduct it.
- After class, review the results and decide what changes, if any, to make.
- Let your students know what you learned from the CAT and how you will use this information.



## Classroom Assessment Techniques (CATS)

- **Why should I use CATs?**
- **For instructors**, more frequent use of CATs can:
- Provide short-term feedback about the day -to-day learning and teaching process at a time when it is still possible to make mid-course corrections.
- Provide useful information about student learning with a much lower investment of time compared to tests, papers, and other traditional means of learning assessment.
- Help to foster good rapport with students and increase the efficacy of teaching and learning.
- Encourage the view that teaching is a formative process that evolves over time with feedback.



## Classroom Assessment Techniques (CATS)

**For students**, more frequent use of CATs can:

- Help them become better monitors of their own learning.
- Help break down feelings of anonymity, especially in larger courses.
- Point out the need to alter study skills.
- Provide concrete evidence that the instructor cares about learning.



## Classroom Assessment Techniques (CATS) - Minute paper

- ***Description:*** During the last few minutes of the class period, ask students to answer on a half-sheet of paper: "What is the most important point you learned today?"; and, "What point remains least clear to you?" The purpose is to elicit data about students' comprehension of a particular class session.
- ***What to do with the data:*** Review responses and note any useful comments. During the following class periods emphasize the issues illuminated by your students' comments.



## Classroom Assessment Techniques (CATS) - The Muddiest Point

- ***Description:*** The muddiest point is an adaptation of the minute paper and is used to find out what students are unclear about. At the end of a lecture or class session, students are asked to write brief answers to the following question: what was the muddiest point in my lecture today? The purpose is to elicit data about students' comprehension of a particular class session.
- ***What to do with the data:*** Review responses and note specific difficulties from students. During the following class periods give a pre-class exercise problem, and re-explain the concepts.
- ***Time required:*** Prep: Low, In class: Low, Analysis: Low



## Classroom Assessment Techniques (CATS) - Chain Notes

- ***Description:*** Students pass around an envelope on which the teacher has written one question about the class. When the envelope reaches a student he/she spends a moment to respond to the question and then places the response in the envelope.
- ***What to Do with the Data:*** Go through the student responses and determine the best criteria for categorizing the data with the goal of detecting response patterns. Discussing the patterns of responses with students can lead to better

## Classroom Assessment Techniques (CATS) - Directed paraphrasing

- ***Description:*** Ask students to write a layman's "translation" of something they have just learned — geared to a specified individual or audience— to assess their ability to comprehend and transfer concepts.
- ***What to Do With the Data:*** Categorize student responses according to characteristics you feel are important. Analyze the responses both within and across categories, noting ways you could address student needs.



## Classroom Assessment Techniques (CATS) - One-sentence summary

- ***Description:*** Students summarize knowledge of a topic by constructing a single sentence that answers the questions "Who does what to whom, when, where, how, and why?" The purpose is to require students to select only the defining features of an idea.
- ***What to Do With Data:*** Evaluate the quality of each summary quickly and holistically. Note whether students have identified the essential concepts of the class topic and their interrelationships. Share your observations with



## Classroom Assessment Techniques (CATS) - Application cards

- ***Description:*** After teaching about an important theory, principle, or procedure, ask students to write down at least one real-world application for what they have just learned to determine how well they can transfer their learning.
- ***What to Do With Data:*** Quickly read once through the applications and categorize them according to their quality. Pick out a broad range of examples and present them to the class.
- ***Time Required:*** Prep: Low, In class: Low,



## Classroom Assessment Techniques (CATS) - Student- generated test questions

- ***Description:*** Allow students to write test questions and model answers for specified topics, in a format consistent with course exams. This will give students the opportunity to evaluate the course topics, reflect on what they understand, and what are good test items.
- ***What to Do With Data:*** Make a rough tally of the questions your students propose and the topics that they cover. Evaluate the questions and use the goods ones as prompts for discussion. You may also want to revise the questions and use them on



## Classroom Assessment Techniques - Misconception/Preconception Check

- ***Description:*** The Misconception/Preconception Check focuses on uncovering prior knowledge or beliefs that may hinder or block further learning. There are many misconceptions in statistics and probability. A quick check of a misconception before the instruction can bring students attention during the instruction.
- ***What to Do With Data:*** Summarizes major misconceptions students made. They should be emphasized during the instruction to help students correct their misconceptions.



## Activity
## Classroom Assessment for your

**In your Introductory Statistics course, you have just completed introducing, stem & leaf graphical displays, numerical summaries, Empirical rule and the relationship between distribution and center measurements. Now you would like to give students a CATS assessment.**

**(1) Write down what you would like to learn from students about their understanding of some of the important concepts you value the most.**

**(2) Determine which CATS tool will serve your purpose.**

## Course Level Assessment Authentic Assessment

- Authentic assessment refers to "Assessment activities that represent literacy behavior of the community and workplace and that reflect the actual learning and instructional activities of the classroom and out-of-school-worlds."



## Common Characteristics of Authentic Assessment

- Asks students to perform, create or produce something
- Encourages student self-reflection
- Measures outcomes of significance
- Taps higher-level thinking and problem-solving skills
- Uses tasks that represent meaningful instructional activities
- Invokes real-world applications
- Uses human judgment (rather than machines) for scoring
- Requires new instructional and assessment roles for teachers
- Provides self-assessment opportunities for students
- Provides opportunities for both individual and group work
- Encourages students to continue the learning activity beyond the scope of the assignment
- Defines explicit performance criteria



## An Authentic Assessment Technique: Performance-based Assessment

- Performance assessment refers to assessment activities in which students are required to demonstrate their level of competence or knowledge by creating a product or response scored so as to capture not just the "right answer", but also the reasonableness of the procedure used to carry out the task or solve the problem.
- To "act upon and bring to completion." Involves displaying one's knowledge effectively to bring to fruition a complex product or event. Performance assessments typically involve the creation of products.

(Kulieke, et al., 1990)



## Performance-based assessment (Sweet, 1989)

**Two Components:**

- A clearly defined task.
- A list of explicit criteria for assessing performance.

**Four Assumptions:**

- ***Knowledge is Constructed:*** Research tells us that students show greater interest and perform at higher levels of learning when they are required to organize facts around major concepts and then actively construct their own understanding of those concepts. They also retain knowledge better.

  Active participation is the key to all performance assessments.

- ***The Task is Worthwhile:*** The ideal performance task is inherently instructional, actively engaging students in



## Performance-based assessment

### Four Assumptions (Continued)

- ***Better Assessments Improve Teaching:***

  Assessment's overall purpose is "to provide valid information for decision making." (Kulieke, et al., )

  When teachers prepare students for a performance task, they must carefully describe the task and the standards that will be used to evaluate performance. When teachers are informed of the learning progress and difficulties of their students they can then make better decisions about content and instruction.

- ***Meeting Criteria Improves Learning:*** Students should be active participants in their own learning. They perform better when they know what goals they are



## Criteria for Selecting Performance Assessment Tasks

- Does the task truly match the outcome(s) you're trying to measure?
- Does the task require the students to use critical thinking skills?
- Is the task a worthwhile use of instructional time?
- Does the assessment use engaging tasks from the "real world?"
- Can the task be used to measure several outcomes at once?
- Are the tasks fair and free from bias?
- Will the task be credible?
- Is the task feasible?
- Is the task clearly defined?

## Activity
## Task Determination – A Case for Intro Stat

A common misconception about the representativeness of sample: People tend to think if a survey will be taken in two different sizes of populations, the sample sizes should also be proportional to the population.

Write a task to assess students' understanding of the representativeness of sample.



## Turning Course Projects and Examines into Student Learning Outcomes Information required at the Program Level

- Student learning outcomes assessment does not need to be an exceptionally additional workload. We have been conducting some sort of learning outcomes assessment for INDIVIDUAL student and his/her grade.
- A difference between assigning grades and what we called 'learning outcomes assessment' is about how we organize the assessment results, and how we use the assessment results.
- A course level learning outcomes assessment should be integrated into regular class projects and/or examines.
- It takes only one extra step when we assess each student's learning outcomes using projects, examines



## Formative Vs. Summative Assessment

- ***FORMATIVE EVALUATION*** is aimed at personal teaching improvement; it is designed to provide an instructor with information he/she can use in current and future classes. This is about learning outcomes/processes assessment.
- ***SUMMATIVE EVALUATION*** is an after-the-fact assessment of a course. End of semester evaluations. This is about final grades.



## Grades Vs. Assessment

**Grade** is a tool for assessing **individual** student's achievement based on individual instructor's grading standards. It is conducted at course level.

**Assessment** is a tool for assessing a **group** of students' **progress as well as outcomes** based on the learning goals/objectives that can be performed at:

- institutional level: eg., overarching educational goals
- college level: eg., alumni performance,
- department level: eg., placement,
- program level: eg., capstone experience,
- course level: eg., class performance of key concepts, student



## An Example: Course Level assessments Vs. Grades – Primary Trait Analysis

Instructors A and B both teach Introductory Statistics

They cover the same topics, test the same concepts using different HWs and tests. Here are the report cards from their final exams:

| | Instructor A | | | | Instructor B | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Grade | Key Concepts | | | | Grade | Key Concepts | | |
| ID | final | Key#1 | #2 . . . #10 | | ID | Final | #1 | #2 ... | #10 |
| 1 | A | 3/5 | 4/6 . . . 4/4 | | 1 | B | 4/7 | 3/3… | 3/5 |
| 2 | B+ | 2/5 | 3/6 . . . 3/4 | | 2 | A- | 6/7 | 2/3 | 4/5 |
| . | . | | | | . | . | | | |
| 40 | C | 2/5 | 2/6 . . . 2/4 | | 30 | D | 2/7 | 0/3 | 2/5 |



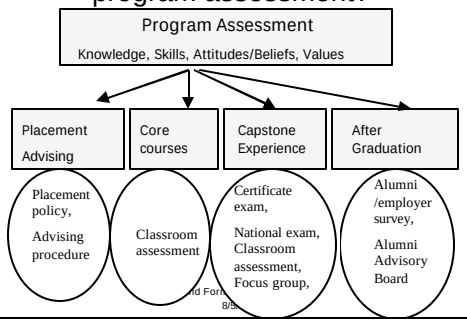
## Is this assessment useful?

- Students have difficulty with Key concept #1
- Students can apply concept # 10 generally well.
- Further analysis of the questions for key concept #1 indicates more than 60% of students can deviate the difference between equally likely and unequally likely.
- A follow up action is taken by both instructors by developing an activity that stresses this difference of these two concepts.

## Primary Trait Analysis

- The key concepts are the primary traits that are identified to meet the learning goals/objectives.
- This is a way to integrate grading with assessment (One stone, two birds).
- Although different instructors may use completely different grading system, as long as the primary key concepts (primary traits) are the same for a given course, the above primary trait analysis can be applied.
- The summarized results from these key concepts (primary traits) can be combined into a useful summary for discussion among faculty, and used

## How can the key concepts analysis (Primary Trait Analysis) be used for program assessment?






## Turning Course Assessment into Program Assessment Activities

- Use Curriculum Alignment techniques

| Program Objectives | Related Courses | Core courses for assessment | Course Objectives | Primary Traits |
|---|---|---|---|---|
| K1-K3 | STA382 | STA 580/575 | Obj#1-5 | Project Rubric scores |
| S1-S2 | STA 580/575 | | Final Exam | Key Concepts |
| K4-K5 | STA 580/575 | STA 590 | Obj#1-6 | Final Exam Key concepts |
| S1-S4 | STA 590 | | Project | Rubric scores |
| K2,K4 | STA 584 | STA 585 | Obj#1-4 | Final Exam Key concepts |
| | STA 585 | | | |



## Turning assessment activities into a classroom-based research

- Through conducting assessment activities, teachers are always engaging in investigation and striving for improved learning. This is a type of 'action research'. The key to action research is to pose a question or goal, and then design actions and evaluate progress in a systematic, cyclical fashion as the means are carried out. Below are four major ways that you can become involved as an action researcher.

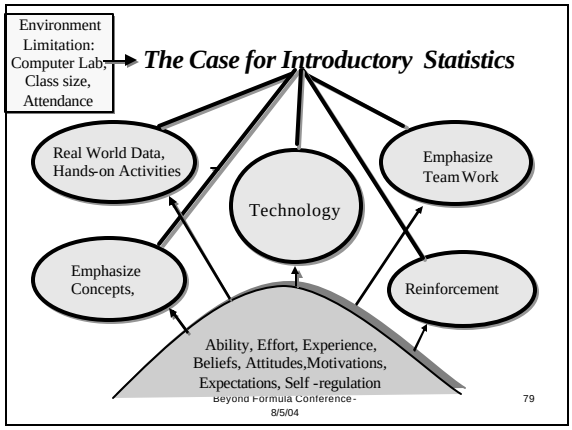


## Turning assessment activities into a classroom-based research

- A classroom assessment activity is usually planed to investigate a certain hypothesis that is related to learning goals.
- Once the activity is conducted for several semesters and that the results are feedback to students and are used for adjusting different teaching strategies, the longitudinal activity is actually a classroom-based research activity.
- The results and lessons learned from this longitudinal assessment activity can be turned into a research work for publication.



## The Case of Introductory to Statistics






## A Call for Change – Workshop on Statistics Education Organized by Hogg in 1990 (MAA Notes # 26. 1992)

- **Problems in Statistical Education**
  - Statistics teaching are often stagnant with little enthusiasm, teachers resist changes (Teacher attitudes/beliefs)
  - Techniques are often taught in isolation, with little consideration of motivations. (Teaching strategy)
  - Statistics are too often presented as a bunch of mathematics (Teaching strategy)
  - Many teachers do not take the advantage of technology (Teaching strategy)
  - Many teachers may not have an adequate backgrounds in knowledge (Content knowledge)
  - Students often view statistics as 'hard' & 'boring subject, and have a fear of mathematical reasoning (Student attitudes/beliefs)
  - Statistics notations is unnecessarily complex and inconsistent (Content difficulty)



## Recommendations from an ASA/MAA Statistics Education Focus

- **Cobb (1992, in the MAA Notes #22)**
  - **EMPHASIZE STATISTICAL THINKING** especially on
    - The need for data
    - The importance of data production
    - The omnipresence of variability
    - The quantification and explanation of variability
  - **MORE DATA & CONCEPTS: LESS THEORY, FEWER RECIPES**
    - More on broad concepts, principles and interpretations in real context
  - **FOSTER ACTIVE LEARNING**
    - Creating a learning environment that encourages active engagement, increase motivation and develops a sense of learning community
    - Alternative to lecturing may include Group activities, Labs, In-class hands-on activities, Written & oral presentations, Projects, and so on

Moore (1997) 'The New Pedagogy and New Content' discussed the importance of statistical thinking and data and provided general suggestions on how to teach the 'new content' using technology and different teaching strategies.



## Six Recommendations for Introductory Statistics Courses – 2004 Draft (Guidelines for Assessment and Instruction in Statistics Education (GAISE) project, Led by Joan Garfield)

- Emphasize statistical literacy and develop statistical thinking
- Use real data
- Stress conceptual understanding rather than mere knowledge of procedures
- Foster active learning
- Use technology to develop conceptual understanding and analyze data
- Use assessments to improve and evaluate



## How do students learn?

(a) Students learn by constructing knowledge [Cognition]

(b) Students learn by active involvement in learning activities [Cognition]

(c) Students learn to do well only what they practice doing [Cognition]

(d) Students learn better if they receive consistent feedback [Progress Assessment – Learning goals]

(e) Students learn to value what they know will be accessed [Performance Assessment – Learning goals]

(f) Students learn better when learning activities are closely related to their personal experience. [Affect – Experience, attitudes]

(g) Students learn better when learning goals are clearly articulated and emphasized throughout the semester. [Metacognition]

(Ref: E.g. Garfield (1995), Gordon (1995), Lee (2000), Steven Von Glasersfeld 1987, Vygotsky 1987, Leont'ev 1981 and so on.

(h) Students learn better when the instructor cares for their



## Course Level Assessment: What should be assessed in Intro to Statistics?

- Misconceptions
- Statistical Literacy
- Statistical Reasoning
- Statistical Thinking

Based on Concepts or integrated set of concepts.

**TO ASSESS WHAT WE VALUE AND WISH STUDENTS TO LEARN.**

## Assessing Misconceptions

- **Random Process Vs. Random Product**
  - Tossing a coin 10 times gives the product: HHHHHHHHHH, In the 11th tossing of the same coin, will H or T be more likely to occur?

    H will be more likely to occur since if one make a prediction based on the given data (product or outcome), the pattern of the product indicates highly likely H should occur.

Possible Misconception

The confusion of Random Process Vs. Random Product. When the process is ignored, purely based on the product, one may conclude that using the product pattern, the prediction of the next tossing is highly likely to be H, and begin to find causes that produce the product such as the coin may not be a fair coin or the individual who tosses the coin may play some tricks, and so on. This

## Gambler's Fallacy **(Kahneman, Slovic & Tversky, 1982)**

- Tossing a coin 10 times gives the product: HHHHHHHHHH, In the 11th tossing of the same coin, will H or T be more likely to occur?
- This is due to the confusion of 'small sample' Vs. The law of Large number' Many similar misconceptions related to Gambler's Fallacy have been reported in the literature.

Laplace (1814): "I have seen men, ardently desirous of having a son, who could learn only with anxiety of the births of boys in the month when they expected to become fathers. Imagining that the ratio of these births to those of girls ought to be the same at the end of each month, they judged that the boys already born would render more



## Spatial/sequence bias in random selection

- Randomly select a number from the following:

  1, 2, 3, 4, 5, 6, 7

- Randomly select a number from the following:

  4, 1, 5, 7, 3, 2, 6



## Inference of Average Number

The average family has three children. It follows the average age child in this town has two siblings.

Or, more precisely,

The average # of children per family is three. It follows that the average number of siblings per child in this town is two.

- Is this a adequate statement?
  - Discuss when this statement is correct.
  - Give an example to show this is not a correct statement



## Confusion of Conditional Probability
## A similar problem of Monty Hall's Three Doors problem

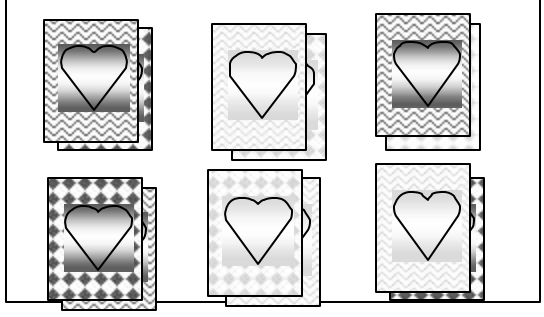
There are three cards: One red on both sides (RR), one White on both sides (WW) and one card has Red on one side and White on the other (RW).

Person A asks B to shuffle the cards out of sight, pick one at random and report the color of (only one side. A can not see the card neither can you.

After B shuffling the cards, picks one and reports "Red". A then offers to bet $10 against $9 that



**Given the card picked "Red", what is the probability the other side of this card is also 'Red'?**



## Expected Value - In a large random sample of men and women, should we expect the men to have more sisters than women? (or women have more brothers than men?)

Work out the case for a family of three children:

| M | F | N | M'B | M'S | F'B | F'S | TMB | TMS | TFB | TFS |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 1 | 2 | 0 | 0 | 0 | 3x2x1 | 0 | 0 | 0 |
| 2 | 1 | 3 | 1 | 1 | 2 | 0 | 2x1x3 | 2x1x3 | 1x2x3 | 0 |
| | | | | | | | | | | |
| | | | | | | | | | | |

## The St. Petersburg Paradox: The Expected Value Vs. Most Likely Outcome

A modern version of the paradox:

Consider a volatile stock that each year, with equal probability, increases in value by 60% or decreases by 40%. What is the $1,000 investment in this stock likely to be worth in 100 years?

By using the expected value, the average change in % per year is 10% gain. Therefore, in 100 years, the stock is expected to be worth of

$$\$1000 \times (1.1)^{100} = \$13,780,612$$

However, given the stock is equally prone to increase by 60% or decrease by 40%, the most probable scenario is increasing for 50 years and decreasing for 50 years.



## Simpson's Paradox

### Graduate Admissions at UC Berkeley (1973)

UC Berkeley was charged with having discriminated against women in their graduate admissions process for the Fall of 1973. The following table is a summary of # of accepted and denied applicants in each of the six largest graduate programs at the time:

| | Men | | | Women | | |
|---|---|---|---|---|---|---|
| | Accepted | Denied | Total | Accepted | Denied | Total |
| A | 511 | 314 | 825 | 89 | 19 | 108 |
| B | 352 | 208 | 560 | 17 | 8 | 25 |
| C | 120 | 205 | 325 | 202 | 391 | 593 |
| D | 137 | 270 | 407 | 132 | 243 | 375 |
| E | 53 | 138 | 191 | 95 | 298 | 393 |
| F | 22 | 351 | 373 | 24 | 317 | 341 |
| Total | 1195 | 1496 | 2691 | 559 | 1376 | 1935 |



## Simpson's Paradox

In the combined Table:

P( Accepted | Women) = 559/1835 = 30.5%

< P( Accepted Overall) = 1754/4516=38.8%

While for each individual program:

Program A: P( Accepted | Women) = 89/108 = 82.4%

> P( Accepted for A) = 600/933 = 64.3%

Program B: P(Accepted|Women) = 17/25 = 68%

> P(Accepted in B) = 369/585 = 63%



## A Small Wager

Jack and Jill are playing the following game:

Tossing a fair coin until the winner is decided.

Game (1) Jack chooses first and bets on the sequence of 3 H's (HHH) and Jill chooses after Jack and bets on THH sequence. Does Jack and Jill have equal chance of winning? What is the odd for Jill to win?

Game (2) Jack this chooses first and bets on THH. Jill chooses after Jack and bets on TTH. Does Jack and Jill have equal chance to win?



## A Small Wager

(Nickerson, 2004. p. 147; Hombas, 1997)

**The probability that Jill will win to play the Triplet game with Jack is given in the following table:**

| | HHH | HHT | HTH | HTT | THH | THT | TTH | TTT |
|---|---|---|---|---|---|---|---|---|
| HHH | | 1/2 | 2/5 | 1/8 | 5/12 | 3/10 | 1/2 | 1/2 |
| HHT | 1/2 | | 2/3 | 2/3 | 1/4 | 5/8 | 1/2 | 7/10 |
| HTH | 3/5 | 1/3 | | 1/2 | 1/2 | 1/2 | 3/8 | 7/12 |
| HTT | 3/5 | 1/3 | 1/2 | | 1/2 | 1/2 | 3/4 | 7/8 |
| THH | 7/8 | 3/4 | 1/2 | 1/2 | | 1/2 | 1/3 | 3/5 |
| THT | 7/12 | 3/8 | 1/2 | 1/2 | 1/2 | | 1/3 | 3/5 |
| TTH | 7/10 | 1/2 | 5/8 | 1/4 | 2/3 | 2/3 | | 1/2 |
| TTT | 1/2 | 3/10 | 3/10 | 1/8 | 2/5 | 2/5 | 1/2 | |

## The King's Folly

There once was a King who worried about the fact men and women were born in about equal numbers. This bothered him because the ideal situation, in his view, was to have more women than men so that he could have more wives. In order to accomplish his goal, he ordered the following decree:

Any family could have as many children as it wished until it produced its first boy, whereupon, the family must cease from any more child.

After several years of the decree, the King was very pleased that there were many families having many girls and no family has more than one boy. His decree must



## The Sample Size Intuitions
**(Kahneman, Slovic & Tversky, 1982)**

In a mid-size city, there are two hospitals. In the large hospital, about 45 babies are born each day, and in the small hospital, 15 babies are born each day. As you know about 50% of all babies are boys. The exact percentage, however, varies from day to day. Sometimes, it may be higher than 50%, sometimes lower.

For a period of one year, each hospital recorded the days on which (more/less) than 50 percent of the babies born were boys. Which hospital do you think recorded more such days? Why?



## Sample Size Intuitions

Of the participants, about half selected the large hospital recorded more of such days.

They ignored the fact that larger deviation from the mean (in percentage sense) are more likely in smaller samples than the large ones.

- Sample size matters, but were either ignored or misunderstood.



## Representativeness of Sample

- A survey will be conducted to estimate the percentage of voters preferring candidate X in two cities:

  City A: population 100,000

  City B: population 500,000

A random sample will be chosen from each city. Since City B is five times as the size of City A, if we take 2000 from City A, we should also take the sample with size proportional to what we choose for City A, which would be about 10,000.



## Perception of Co-variation and Contingency

**Biased Focus:** The cell of a two-way table that tends to get most attention is the one that represents cases in which both of the variables of interest occurred such as the (Plus, Plus) cell. (e.g., Crocker, 1981; Anderson, 1990)

**Expectations and Cause-Effect Assumptions**

Co-variation is especially difficult to detect when it occurs between variables that are not expected to be related (e.g., Peterson, 1980).

When co-variation is identified, people have a tendency to jump into interpreting one variable is



## Cause-Effect or Not

Some examples:

It is found that the math IQ for kindergarteners is highly related to the foot size. Thus, if schools is interested in selecting gifted students from this age group, all they need to do is to measure their foot size.

It is found that the profit of a company is positively correlated with the CEO's salary. Therefore, all the company needs to do is to increase their CEO's salary I order to have profit.

The life expectancy is found highly positively correlated with the average # of TV's per 1000

# Some Unexpected Difficulties using Histograms to Reason Variation

(Lee & Meletiou, 2004)

We tested students with four Tasks designed to investigate students' reasoning about histograms (see Appendix D), and analyzed 162 students' responses to these items through three semesters. We identify four main types of student difficulties in constructing, interpreting and applying histograms in different real world contexts:

1. Perceiving histograms as displays of raw data with a focus on individual cases rather

E.g., When asking students to draw a reasonable histogram for the salary for individuals who are 40 years or older before retirement. Many students draw a histogram similar to the one on the right and interpreted as Y: salary and X: individual case ID.)

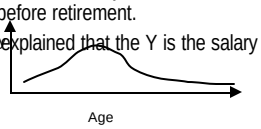


## Some Unexpected Difficulties using Histograms to Reason Variation

2. Tending to interpret histograms as two-variable scatterplots or as time sequence plots.

For the same question about drawing a histogram to summarize the salary of individuals who are 40 years or older before retirement.

Many explained that the Y is the salary and X is the Age.



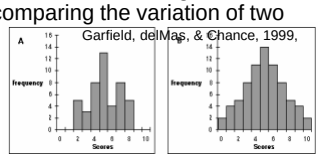



## Some Unexpected Difficulties using Histograms to Reason Variation

3. Tending to look at the vertical axes and compare differences in the heights of the bars when comparing the variation of two histograms.

Garfield, delMas, & Chance, 1999,

Many students think that A has larger variation because it is bumpier.






# Some Unexpected Difficulties using Histograms to Reason Variation

4. Tending to think deterministically when interpreting a distribution in real world contexts.

When mean > median, regardless how large the standard deviation is, many students immediately conclude the shape of the distribution is skewed to right.

E.g., An insurance company is interested in the cholesterol levels of individuals in our community that are 40 years of age or older. A random sample of 100 individuals was chosen from this population and the following information was collected:

Average = 158 (mg), Median = 160 (mg), Standard deviation = 20 (mg)



# Conjunction Fallacy

**(Kahneman, Slovic & Tversky, 1982)**

A tendency of overestimating the probability of the joint occurrence of events:

"Bill is 30 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics, but weak in social studies and humanities."

Rank the following statements about Bill from most to least likely:

(a) Bill is an accountant

(b) Bill plays jazz for a hobby

(c) Bill is an account who plays jazz for a hobby.



# Conjunction Fallacy (?)

The above scenario may have different interpretations:

- When people make a judgment about Bill, the probability concept that P(A and B) can not be larger than P(A) or P(B) was not be used in their decision making.
- In common language, when people read statement (b) and (c ), one might interpret (b) as: Bill is an account who does not play Jazz for a hobby.

Thus, this may be due to the use of the language, not the misconception about P(A and B) has to be no larger than P(A) or P(B).

## Equality Probability Bias

When ask students the probability of tossing a balance die for a large # of times, what would be the chance of each face to occur. There is no problem that P(each face) = 1/6

When asking students " Carl Lee plays golf with Tiger Woods a game", what would the chance of winning for Carl. Little problem about this question: P(Carl wins) is practically zero.

However, when asking students: Carl plays golf with Tiger Woods for three games". Use the notation W stands for Carl Wins a game. The sample space is {WWW, WWL, WLW, LWW, WLL, LWL, LLW, LLL}. What is the P(WWW)?



## Real Context Vs. Statistical Language

Using the same context that Carl plays golf with Tiger Woods three game.

When asking students the same question using the context only: what is the chance that Carl will win all three games? Many student are then able to sense that Carl stands virtually no chance in the every day language, not in terms of statistical language.

For many students, statistical language stays within the world of statistics. When connect with context, people lose the connection between context and statistical reasoning.



## What to Assessment for Intro to Statistics

**Assessing**

- Statistical Literacy
- Statistical Reasoning
- Statistical Thinking

- Recommendations from the GAISE Report (2004 Draft) on assessment
- Performance-based assessment for Intro to Stat
- WebArtist and other resources



## What is Statistical Literacy?

- The GAISE Report (2004, Draft) Defines Statistical Literacy as:

  Understanding the basic language of statistics (e.g., knowing what statistical terms and symbols mean and being able to read statistical graphs), and understanding some fundamental ideas of statistics. For readings on statistical literacy see (More can be found e.g., in Gal (2002), Rumsey (2002), Utts (2003), and Watson (2003)).



## What is Statistical Reasoning?

The way people reason with statistical ideas and make sense of statistical information (including being able to make proper data summary and proper interpretation of the statistical information). Under the reasoning is the conceptual understanding of important ideas such as center, variation, sampling, distribution, randomness, association and the connection among these ideas.

(E.g., Garfield, 2003; Garfield & Gal, 1999; Lovett, 2001)



## What is Statistical Thinking?

GAISE Report (2004 Draft) defines Statistical Thinking as:

The type of thinking that statisticians use when approaching or solving statistical problems. Statistical thinking has been described as understanding the need for data, the importance of data production, the omnipresence of variability, and the quantification and explanation of variability.

(see also Snee, 1990; Cobb 1992; Wild & Pfannkuch,

## An Analogy: The Amazon Rescue Mission A Mission of Statistical Thinking

You are assigned a rescue mission in the Amazon forest, where you have never been. The rescue mission in this case is solving problems by properly using any techniques. This process involves with adventure and a lot of decision making.

**The Rescue Mission is at the level of Statistical Thinking.**



## To Accomplish the Mission: Requires Statistical Literacy

So you must be trained for the mission. Two major areas of training are needed. One is to learn how to use every tool that is available to you. You actually see them, use them and practice until you are very good at operating every tool.

**This is analogous to Statistical Literacy.**



## To Accomplish the Mission: Requires Statistical Reasoning

The second level of the training for the rescue mission is to completely understand the properties of the tools and simulate possible situations of Amazon forest, which is indeed unknown to you. However, because of the characteristics, patterns, and knowledge that are learned from experiences or documented theories about forest, we can imagine the potential situation and simulate that situation, and learn its properties and patterns before going into the forest. This is equally important as the first stage of training.



**This is analogous to the level of Statistical**

## Recommendation 6: Use assessments to improve and evaluate student learning

(GAISE 2004 Report)

- Students will value what you assess.
- Make sure your assessments are aligned with learning goals.
- Make sure assessments focus on understanding key ideas and not just on skills, procedures, and computed answers.
- It should be done with formative assessments used during a course (e.g., quizzes and midterm exams and small projects) as well as with summative evaluations (course grades).
- Useful and timely feedback is essential for assessments to lead to learning.



***Types of assessment:***

## Suggestions for teachers on Assessments (GAISE 2004 Draft Report):

- Integrate assessment as an essential component of the course. Assessment tasks that are well coordinated with what the teacher is doing in class are more effective than tasks that focus on what happened in class two weeks earlier.
- Use a variety of assessment methods to provide a more complete evaluation of student learning.
- Assess statistical literacy using assessments such as interpreting or critiquing articles in the news and graphs in media.
- Assess statistical thinking using assessments such as student projects and open-ended investigative tasks.



## Suggestions for student assessment in large classes (GAISE 2004 Draft Report)

- Use small group projects instead of individual projects.
- Use peer review of projects to provide feedback and improve projects before grading.
- Use multiple choice items that focus on choosing good interpretations of graphs or selecting appropriate statistical procedures.
- Use discussion sections for student presentations.

## Applying Performance-based Assessment to Intro to Statistics

- **Two Components:**
  - A clearly defined task.
  - A list of explicit criteria for assessing performance
- **Four Assumptions:**
  - ***Knowledge is Constructed:*** Active participation is the key to all performance assessments.
  - ***The Task is Worthwhile:*** Performance tasks are open-ended and assess an array of knowledge and skills related to the curriculum.
  - ***Better Assessments Improve Teaching:*** Assess what you value the most, and use the result to improve teaching.
  - ***Meeting Criteria Improves Learning:*** Assess



## Applying Performance-based assessment techniques – Intro to Stat

- **Group projects or individual projects**
  - Evaluating criteria – rubric
- **Presentation**
  - Evaluating criteria – rubric
- **Scenario critique**
  - Evaluating criteria – Key concepts, some times, no correct answer
- **Multiple Choice questions with reasoning explanation**
  - Evaluating criteria - Answer keys for the multiple choice questions and key concepts for the reasoning
- **Open-ended questions focusing on reasoning and problem solving**
  - Evaluating criteria – Key concepts, rubrics



## Assessing Statistical Literacy, Reasoning & Thinking using the same context (Garfield, Chance, 2003; WebArtist)

Problem: The following stem-and-leaf display displays the average annual snowfall amounts (in inches, with the stems being tens and leaves being ones) for a random sample of 25 American cities:

| | |
|---|---|
| 0 | 000000024 |
| 1 | 028 |
| 2 | 00228 |
| 3 | 8 |
| 4 | 2248 |
| 5 | 48 |
| 6 | 0 |



## Three levels of assessment using the Snowfall context (Continued)

- ***Statistical literacy*:**
  - Describe this distribution.
- ***Statistical reasoning:***
  - Without doing any calculations, would you expect the mean of the snowfall amounts to be larger, smaller, or about the same as the median? Why?
- ***Statistical thinking*:**
  - A researcher has data on the average snowfall for these same cities from 20 years ago. She wants to test whether snowfall amounts are higher now than they were 20 years ago. Describe an appropriate



## Activity
## Applying Performance-based Assessment

You notice that students are confused between histogram and scatterplot. Write a task to assess if students have learned the difference, and if not, why they are confused so that you can develop additional activities to help them.

(You may use the three level strategy: Literacy, Reasoning and Thinking)



**Welcome to the ARTIST Website!**

Home Page
News
Events
Roundtable Conference
**Assessment Builder**
Questions and Answers
Resources
Ways to participate
Feedback
Publications and Presentations
ARTIST Glossary
About ARTIST
Participants' Page

**Our goal is to help teachers assess:**
**- Statistical literacy**
**- Statistical reasoning**
**- Statistical thinking**
**in first courses of statistics.**

This website provides a variety of assessment resources for teaching first courses in Statistics. Currently we provide assessment items, and articles

**http://www.gen.umn.edu/artist/**

# Introduce WebArtist

**Web site: http://www.gen.umn.edu/artist/**

1. **Click on 'Assessment Builder'.**
2. **Read through the general direction.**
3. **Click on 'Go To Assessment Builder' at the bottom of the page.**
4. **Take you to the Login page. First time user is required to Register.**
5. **After register, you will receive a login Pass Word.**
6. **On the login page, enter your e-mail and PW.**
7. **Begin to build your test set.**



# Reference


- Angelo, T.A. & Cross, P.K. (1993). *Classroom Assessment Techniques* (2nd ed.). San Francisco: Jossey-Bass.
- APA Division Committee Psychological Principles: Guidelines for teaching of educational psychology in teacher education programs. NEP/15 Newsletter for Educational Psychologists 19(1), November, 4-5, 8.
- Chance, B. (2003). Components of Statistical Thinking and Implications for Instruction and Assessment. Journal of Statistical Education, http://amstat.org/publications/jse/v10n3/chance.html.
- Cobb, G. (1992), "Teaching Statistics," in *Heading the Call for Change: Suggestions for Curricular Action,* eds. L. Steen, Washington, D.C.: Mathematical Association of America, pp. 3-33.
- Davis, B.G. (1993). *Tools for Teaching.* San Francisco: Jossey-Bass.
- delMas, R. (2003). Statistical Literacy, Reasoning, and Learning: A Commentary. Journal of Statistical Education, http://amstat.org/publications/jse/v10n3/delmas_discussion.html.
- Elliott, Stephen N. (1995). Creating Meaningful Performance Assessments.. http://www.ericfacility.net/ericdigests/ed381985.html ERIC Digest E531
- EmTech Emerging Technology at http://www.emtech.net/index.shtml
- GAISE Report (2004). Guidelines for Assessment and Instruction in Statistics Education (GAISE) project (Led by Garfield).





- Gal, I., and Garfield, J. (1997). Curricular Goals and Assessment Challenges in Statistics Education. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education.* Burke, VA: IOS Press.
- Gal, I. & Garfield, J. (Eds.) (1997).*The Assessment Challenge in Statistics Education*. Amsterdam: IOS Press.
- Garfield, J. B. (1994), "Beyond Testing and Grading: Using Assessment to Improve Student Learning," *Journal of Statistical Education* [Online], 2(1). (www.amstat.org/publications/jse/v2n1/garfield.html)
- Garfield, J.B. (1995). How student learn statistics. *International Statistical Review*, 63, 1, 25-34.
- Garfield, J. (2003). The Challenge of Developing Statistical Reasoning. Journal of Statistical Education, http://amstat.org/publications/jse/v10n3/garfield.html.
- Garfield, J. (2003). Assessing Statistical Reasoning. Statistical Research Journal 2(1), 22-38, http://fehps.une.edu.au/serj
- Garfield, J. and Chance, B. (2003). The Web-based ARTIST: Assessment Resource Tools for Improving Statistical Thinking. Presented at the Symposium: Assessment of Statistical Reasoning to Enhance Educational Quality, AREA Annual Meeting, Chicago, April,





- Hartman, H. J. Kluwer Academic Publisher. (2001) (Ed.) Metacognition in Learning and Instruction.
- Hombas, V. C. (1997). Waiting time and expected waiting time – paradoxical situations. The American Statistician, 51, 130-133.
- Jane Ehrenfeld (2004). Alternative Instruction and Alternative, Performance-Based Assessment - An Annotated Bibliography.
- Kahneman, D., Slovic, P. & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases, Cambridge, UK: Cambridge University Press.
- Kulieke, M. J. Bakker, C. Collins, T. Fennimore, C. Fine, J. Herman, B.F. Jones, L. Raack, M.B. Tinzmann (1990). Why Should Assessment Be Based on a Vision of Learning? NCREL, Oak Brook.
- Leahey, T. H. & Harris, R. J. (1997) Learning and Cognition (4th Edition). Prentice Hall.
- Lee, C. & Meletiou-Marotheris, M. (2004). A missing link toward understanding variability: The unexpected difficulty of histograms. Submitted.
- Lovett.. (2001). A collaborative convergence on studying reasoning processes: A case study in statistics. Cognition and Instruction: Twenty-Five Years of Progress, eds. D. Klahr and S. Carver, Mahwah, NJ:





- Moore, T. (2001) (Eds.) Teaching Statistics: Resources for Undergraduate Instructors, Washington, D.C>: Mathematical Association of America and American Statistical Association.
- Nickerson, R. S. (2004). Cognition and Chance – The Psychology of Probabilistic Reasoning. Rubric Generators at http://www.teach-nology.com/web_tools/rubrics/
- Schroeder, C. (1993). *New Students – New learning styles*. Collected in *Learning from Change: Landmarks in Teaching and Learning in Higher Education from Change Magazine, 1969-1999* (Ed., Dezure, D., 2000), Published by the *American Association for Higher Education*, pp. 83-84.
- Sims,. Rand Sims, S. (1995) (Eds.) The Importance of Learning Styles. Lawrence Erlbum Assoc. Inc..
- Snee, R. D. (1990). Statistical Thinking and Its Contribution to Total Quality. The American Statistician, 44, 116-121.
- Sternberg, R. J. (1999) (Ed.). The Nature of Cognition. The MIT Press.
- Sweet, D. (1989) *A Question of Thinking: A First Look at Students' Performance on Open-ended Questions in Mathematics*, copyright 1989, California Department of Education, P.O. Box 271, Sacramento, CA 95812-0271. http://www.ed.gov/pubs/OR/ConsumerGuides/perfasse.html (Reprint)
- Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, Vol 67 (3), 223-265.




# Appendix

## An example – Performance-based assessment for the representativeness of sample

A survey will be conducted to estimate the percentage of voters preferring candidate X in two cities:

City A: population 100,000

City B: population 500,000

A random sample will be chosen from each city. Since City B is five times as the size of City A, if we take 2000 from City A, we should also take the sample with size proportional to what we choose for City A which would be about 10,000.



## An example – Performance-based assessment for "histogram Vs. scatterplot"

A study is designed to investigate the relationship between age and their salary for people who are 40 years or older, yet to retire. 100 individuals are selected for the study.

(a) In analyzing a data set such as this, it is a good idea to look at the summary of the data. Determine and explain what plot is appropriate to demonstrate the distribution of the salary of these 100 individuals who are 40 years or older.

Explain what is on the Y-axis and what's on the X-axis for this plot.

Describe the most likely shape of the distribution of this salary variable and explain your reason.

(b) Determine and explain what kind of is appropriate to demonstrate the relationship between salary and age.



## Answer to Expected Value - In a large random sample of men and women, should we expect the men to have more sisters than women? (or women have more brothers than men?)

Work out the case for a family of three children:

| M | F | N | M'B | M'S | F'B | F'S | TMB | TMS | TFB | TFS |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 1 | 2 | 0 | 0 | 0 | 3x2x1 | 0 | 0 | 0 |
| 2 | 1 | 3 | 1 | 1 | 2 | 0 | 2x1x3 | 2x1x3 | 1x2x3 | 0 |
| 1 | 2 | 3 | 0 | 2 | 1 | 1 | 0 | 1x2x3 | 2x1x3 | 2x1x3 |

## Motivation for Studying the Difficulties using Histogram for Reasoning Variation

- *When constructing a histogram for describing the distribution of salary for individuals that are 40 years or older and have not yet retired:*
  - *a) What goes on the vertical axis Y?*
  - *b) What goes on the horizontal axis X?*
- Most students in the study confused the histogram with a scatterplot of salary vs. age.
- They thought that *'the graph is skewed-to-the right because as people approach retirement, their salary gradually drops.*



## Study and participants

- Students were given four tasks related to the construction, interpretation and application of histograms in real world scenarios. These four tasks were part of the test1 and/or final exam.
- The study site was an introductory statistics course at Central Michigan university.
- 162 students participated in three-semester study.
- 75% of the students were Business majors, 25% were other non-science majors.
- The prerequisite was College Algebra. Few students had taken Precalculus level or higher courses.
- Approximately 55% of the students were female.



**Task A (Given in Test 1, Descriptive stat and Probability)**

An insurance company is interested in the cholesterol levels of individuals in our community that are 40 years of age or older. A random sample of 100 individuals was chosen from this population and the following information was collected:

Sample size = 100 Average cholesterol level = 158 (mg)
Median cholesterol level = 160 (mg) Standard deviation = 20 (mg)

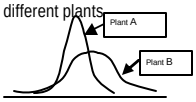
Q1: Based on the information above, the shape of the distribution of cholesterol levels for individuals in the community of age 40 or more is more likely to be _____.
Explain the reason:

Q2: When constructing a histogram for the cholesterol level data
What goes on the horizontal axis *X*?_____
What goes on the vertical axis *Y*?_____

Q3: An individual has a cholesterol level of 188 mg. Is this an unusually high cholesterol level? Why or why not?

## Task B (given in Test 1)

The following graph shows the distribution of the width of window frames manufactured at two different plants.




Which of the following statements is correct? (Choose one).

1. Width distribution of frames manufactured at Plant A is skewed.
2. Width distribution of frames manufactured at Plant A has a similar variation to the width distribution of frames manufactured at Plant B.
3. Width distribution of frames manufactured at Plant A has a larger variation than width distribution of frames manufactured at Plant B.
4. Width distribution of frames manufactured at Plant A has a smaller variation than width distribution of frames manufactured at Plant B.

Explain the reason for your choice.

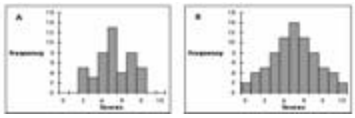


## Task C: (taken from Garfield, delMas, & Chance, 1999, given in Final)

Which of the following distributions show more variability? (Check one):

(a) A has more variability (b) B has more variability

"Choosing distribution with more variability" task



Now, check the statement or statements that led you to select the choice above:

1. Because it is bumpier
2. Because it is more spread out
3. Because it has a larger number of different scores
4. Because the values differ more from the center
5. Other (please explain)



## Task D (Given in Final):

When constructing a histogram for describing the distribution of salaries for individuals that are 40 or older and have not yet retired,

Q1: Explain:

1. What goes on the vertical axis *Y*?
2. What goes on the horizontal axis *X*?

Q2: What would the shape of the salary distribution more likely be? Explain why.