

# Evaluation of Teaching at Stockton

Bill Reynolds, PhD, LCSW

(Based on materials prepared by Heather McGovern)

Associate Professor of Social Work

Director, Institute for Faculty Development

August 2013

# Overview

1. How to know what to do, when, for evaluation of teaching at Stockton
2. The role of different items in evaluation of teaching at Stockton
3. Student evaluations:
  - a. Choosing learning objectives and disciplinary codes
  - b. Interpreting data
  - c. Making comparisons
  - d. Improving teaching and evaluations
4. Peer observations of teaching
5. Teaching portfolio and faculty plan
6. References

# You will provide multiple ways for your teaching to be evaluated

**Stockton policy requires** (pp. 6 – 8, *Procedure for Evaluation of Faculty and Library Faculty*):

- ❑ A self-evaluation of teaching:
  - ❑ Explain one's pedagogical approach (i.e., teaching philosophy) and how one's courses fit into the program and college curriculum
  - ❑ Articulate how one's pedagogy attempts to meet the learning goals in one's courses
  - ❑ Assess how well those goals have been met
- ❑ Documentation of achievement in teaching:
  - ❑ Student evaluations of teaching and preceptorial teaching
  - ❑ Peer evaluations of teaching
  - ❑ Other evidence of effectiveness in teaching

**WHAT TO DO, WHEN**



# Evaluate nearly all courses

- Untenured faculty will use formal evaluations of teaching in all their classes of 5 or more students
- You may have any class observed at any time
- The IFD website has a [list of faculty](#) who have been trained to conduct peer observations, but you are not obligated to use them.

# You need to do some things now.

- Go online through the Stockton portal and identify which of a list of learning objectives are relevant to your class. Your deadline for doing this is the 2<sup>nd</sup> precepting day, **Nov. 6**. Directions will follow later.
- Allow time on syllabi for evaluations from Dec. 2- Dec. 11 (if doing them in class).
  - Default is online administration, which doesn't require class time

# One more thing I recommend you do now: Plan an observation

- You are not required to have any observations completed in your first semester, but I **strongly recommend** that you have one.
- Benefits
  - Demonstration of initiative
  - Lower expectations
  - Gives you something to write about in first year file, which is due Jan. 10, 2014—making it, in reality, a first semester file.

# STUDENT EVALUATIONS

What is your perspective on the value of student evaluations of your teaching?

# Student evaluations are one part of evaluation of teaching.

The IDEA Center “strongly recommends that additional sources of evidence be used when teaching is evaluated and that student ratings constitute only 30% to 50% of the overall evaluation of teaching.” Primary reasons:

- “some components of effective teaching are best judged by peers and not students”
- “it is always useful to triangulate information...”
- no instrument is fully valid
- no instrument is fully reliable

# Quick, simplified definitions...no insult intended!

- **Validity:** the tool measures what it says it measures
- **Reliability:** measures will tend to be consistent, repeatable
- **Representative:** the people responding would be likely to be typical of the whole group

# Validity: Student teaching evaluations correlate with others' ratings

Administrator .39 to .62

Colleagues .48 to .69

Alumni .40 to .70

Trained observers .50 to .76

Student comments .75 to .93



# How Stockton defines “excellence in teaching” and what students rate

- “A thorough and current command of the subject matter, teaching techniques and methodologies of the discipline one teaches
- Sound course design and delivery in all teaching assignments...as evident in clear learning goals and expectations, content reflecting the best available scholarship or artistic practices, and teaching techniques aimed at student learning
- The ability to organize course material and to communicate this information effectively. The development of a comprehensive syllabus for each course taught, including expectations, grading and attendance policies, and the timely provision of copies to students.
- ...respect for students as members of the Stockton academic community, the effective response to student questions, and the timely evaluation of and feedback to students.”
- “Where appropriate, additional measures of teaching excellence are:
  - Ability to use technology in teaching
  - The capacity to relate the subject matter to other fields of knowledge
  - Seeking opportunities outside the classroom to enhance student learning of the subject matter”

# Reliability and representativeness: Number of classes needed for evaluation

- The IDEA Center “recommends using six to eight classes, not necessarily all from the same academic year, that are representative of all of an instructor’s teaching responsibilities.”
- In a person’s first few years at Stockton, evaluators will not be able to do what is ideal. **This makes having teaching observations and other evidence of good teaching in your file even more important.**

# RELIABILITY

- The number of student respondents affects reliability.
- IDEA reports the following median rates:
  - 10 respondents--.69 reliability
  - 15 respondents--.83 reliability
  - 20 respondents--.83 reliability
  - 30 respondents--.88 reliability
  - 40 respondents.--91 reliability
- **Reliability ratings below .70 are typically suspect.**
- **Starting in Fall 2010, to respond to this issue, Stockton began using a small class instrument (that gathers qualitative rather than quantitative data) for classes of fewer than 15 students.**

# The number of student responders affects representativeness

- Higher response rates provide more representative data. Lower response rates provide less representative data.
- This is especially an area of concern for classes using the online IDEA, which has a lower response rate.
- But any class (even a small one) can have a low response rate, and if this happens in your class you should note it for file readers and note it yourself.

# Recent online response rates

- Range
  - Low: 48.87%, spring 2012
  - High: 69.84%, fall 2012
- Mean: 59.6%
- Explanation
  - During two semesters there was a very short (one week) window during which students could complete online IDEA
  - Some faculty may not have realized their courses defaulted to online administration and failed to remind students
- Mean response rate of all online courses in IDEA database: 67%

# Selecting Objectives and Checking Your Disciplinary Code

- Look at the faculty information form so you can see the objectives from which you can choose...
- I'll project a sample IDEA report.

# The objectives you choose matter.

- Item A. on P. 1 and column 1 in the graph report Progress on Relevant Objectives.
- The “Summary Evaluation” provided on page one of the IDEA report weights *Progress toward Relevant Objectives* at 50% and *Excellent Teacher* and *Excellent Course* at 25%.
- Data on page two reports student ratings on only the items you selected.
- Data on page four reports all ratings.
- On the small class form, students will list objectives they feel they progressed on.

# It matters whether you say something is “important” or “essential.”

- In the IDEA Progress toward Relevant Objectives scores on page one of the report, items of minor importance do not count at all, and items that are “essential” count double items that are “important.”
- Your choice also sends a signal about your philosophy of teaching for the class.



# Which objectives should I select?

Those that make sense for your class!

For example, if you teach a writing class, logically the objective about progress developing communication skills should be “essential.”

# How can you know whether something is important enough to select?

- Pedagogy and assignments **help students progress** on the objective
- A significant portion of a **student's grade** is comprised of a meaningful measurement of how well students have achieved that objective

# Consider program requests

- Your **program** may have some suggestions or requests about what you select (e.g., the writing program and first year seminars suggest objectives for W and first year seminar classes, and science lab courses often have a supervisor who suggests objectives).
- Programs cannot force selections upon you, but they may have good reasons for requesting that people make similar choices, and/or program guidance can make selecting objectives easier for you.

# How many objectives should I select?

- Usually, 3 – 5.
- You should select **at least one** objective as either “important” or “essential,” but selecting too many is often problematic.
  - It is harder for students to make progress if the class has many objectives.
  - Research indicates that student ratings tend to decrease when larger numbers of objectives are selected.

# Myths about objectives

- *I have to choose a certain number.* No. If you choose none as important or essential, then by default all will be important, which is not good. You may select as few as one or as many as all of them. Best practice is to choose 3-5.
- *I have to have at least one essential or I have to have at least one important.* No. You can have any combination of important and essential objectives. Note: if all your selections are essential or important, they'll be equally weighted and it won't matter mathematically whether they are essential or important.
- *It doesn't matter what I select.*

# Activity

- Briefly describe a class you will teach in this fall to a partner.
- Discuss which objectives you would choose and why.

# Selecting a disciplinary code

- Ideally, your code is as good a match to your class as possible. A match has been selected for you. If it is good, do nothing. If you think it could be better, contact me about a possible new one.

# Meaning of IDEA results

- Student evaluations give you students' *perceptions*, which are not always the same as student learning or as reality.
- Assessment of student learning requires direct measurement



# Outlier can affect mean scores.

- IDEA reports mean scores, which can be affected by outliers. Careful evaluators will check the statistical detail on page 4.
- Standard deviations of .7 are typical. Standard deviations of **over 1.2** indicate “unusual diversity.” If the distribution is bimodal, then the class may have contained “two types of students who are so distinctive that what “works” for one group will not for the other. For example, one group may have an appropriate background for the course while the other may be under-prepared....” (IDEA)
- They recommended detailed item examination; there may be issues beyond instructor control.

# Scores and comments can be affected by the halo effect

- Ranters and Ravers, or the halo effect:
  - “the tendency of raters to form a general opinion of the person being rated and then let that opinion color all specific ratings. If the general impression is favorable, the ‘halo effect’ is positive and the individual receives higher ratings on many items than a more objective evaluation would justify. The ‘halo effect’ can also be negative; an unfavorable general impression will lead to low marks ‘across the board’, even in areas where performance is strong.”

# How can you know?

- Look at the pattern of student responses on page 4 or on the student forms. If a form gives someone a 5 all the way down, regardless of whether a class covered a particular learning objective—halo effect!
- In most cases, also true with a 1 or any other number all the way down...

# The Error of Central Tendency can affect scores

- “Most people have a tendency to avoid the extremes (very high and very low) in making ratings. As a result, ratings tend to pile up more toward the middle of the rating scale than might be justified. In many cases, ratings which are "somewhat below average" or "somewhat above average" may represent subdued estimates of an individual's status because of the "Error of Central Tendency.”

# Things evaluators should check

- ❑ The teacher selected objectives. If not, by default, all will be considered “important.” Much information on the first page of the report is worthless.
- ❑ The objectives the teacher chose seem reasonable for the course.
- ❑ The teacher discusses problematic objective choices or irregularities in the class.

# You can help evaluators by noting if you...

- ❑ forgot to select objectives, which seriously impacts the results.
- ❑ later see that you chose objectives poorly.
- ❑ were using objectives in common with a larger group of courses, but those were problematic for your class.
- ❑ need to report an unusual situation that likely affected student progress towards objectives or student perception of the class.

# IDEA compares class results to three groups (page one and two)

- 1) Three years of IDEA student ratings at all IDEA institutions
- 2) Classes at your institution in the most recent five years (excluding classes with no objectives selected, including classes of all sizes)
- 3) Classes in the same discipline in the most recent five years where at least 400 classes with the same disciplinary code were rated

# The validity of comparisons varies

- The validity of comparisons depends on a number of factors, including how “typical” a class is, compared to classes at Stockton or all classes in the IDEA database or how well the class aligns with other classes with the same IDEA disciplinary code.
- Some classes at Stockton align poorly with “typical” classes—say, a fieldwork class or a class with a cutting-edge format.



# External factors can affect comparisons and ratings

- Students in **required courses** tend to report **lower**.
- Students in **lower level classes** tend to report **lower**.
- **Arts and humanities > social science > math** (this may be because of differences in teaching quality or due to quantitative nature of courses, both, or other factors).
- Gender/age/race/culture/height/physical attractiveness and more may be factors, as they are in many other areas of life.
- If the students are told the evaluation will be used in personnel decisions the scores are higher.
- If the instructor is present during the evaluation the scores are higher.

# Some external factors don't usually affect ratings

- Time of day of the course
- Time in the term in which evaluations are given (after midterm)
- Age of student
- Level of student
- Student GPA

# Some disciplinary comparisons are suspect

- Many classes align poorly with disciplinary codes: CRIM stats here, which is compared either with Criminal Justice or with Mathematics. Or developmental writing here, which is higher level than many but also for credit. Or most of our G courses, perhaps particularly our GIS courses.

# We should use converted scores when making comparisons

- IDEA states that “Institutions that want to make judgments about teaching effectiveness on a comparative basis should use **converted scores.**”
- **Converted scores** are reported in the graph and lower table on page one and on page two.

# Why we should use converted scores

- The 5-point averages of progress ratings on “Essential” or “Important” objectives vary across objective. For instance, the average for “gaining factual knowledge” is 4.00, while that for “gaining a broader understanding and appreciation for intellectual/cultural activity is 3.69.
- Unconverted averages disadvantage “broad liberal education” objectives.
- Using converted averages “ensures that instructors choosing objectives where average progress ratings are relatively low will not be penalized for choosing objectives that are particularly challenging or that address complex cognitive skills.”

# Why we should use adjusted averages in most cases

- **Adjusted scores** adjust for “student motivation, student work habits, class size, course difficulty, and student effort. *Therefore, in most circumstances, the IDEA Center recommends using adjusted scores.*”

# How are they adjusted?

“Work Habits (mean of Item 43, *As a rule, I put forth more effort than other students on academic work*) is generally the most potent predictor... Unless ratings are adjusted, the instructors of such classes would have an unfair advantage over colleagues with less dedicated students.”

# How are they adjusted, part II

- “Course Motivation (mean of Item 39, *I really wanted to take this course regardless of who taught it*) is the **second most potent predictor**. . . .unless ratings are adjusted, the instructors of such classes would have an unfair advantage over colleagues with less motivated students.”
- Others (details on [IDEA website](#))



# A critical exception to using adjusted scores

- *“We recommend using the unadjusted score if the average progress rating is high (for example, 4.2 or higher).”*
- In these cases, students are so motivated and hard-working that the teacher has little opportunity to influence their progress, but “instructors should not be penalized for having success with a class of highly motivated students with good work habits.”

# False assumptions about IDEA

- Effective teaching = students make progress on all 12 learning objectives
- Effective teachers = teachers who employ all 20 teaching methods

# Norming sorts people into broad categories

- Scores are **normed**. Therefore, it is unrealistic to expect most people to score above the similar range.
- Statistically, 40% of people **ALWAYS** score in the similar range and 30% above and 30% below that range.

# More thoughts on norming...

- Many teachers teach well. Therefore, the comparative standard is relatively high. Being “similar” is not bad. It is fine.
- If we made a list of 10 teachers at random at Stockton, we’d expect that one would fall into the “much lower” range, two into “lower,” four into “similar,” two into “higher,” and one into “much higher” if we think Stockton teachers are basically comparable to the teachers in the IDEA database (which they tend to be).

# TEACHING OBSERVATIONS

What is your perspective on the value of peer observations of your teaching?

# Teaching observations can...

- Provide faculty with **formative feedback** about their teaching practices.
- Provide faculty with **evaluative feedback** about their teaching practices that they can use in their files as part of their evidence of good teaching.
- In worst case scenarios, **protect students** from poor teaching.

# You are required to have two observations per year.

- As a tenure-track faculty member, you need **two observations** per year, **completed before file deadlines** (except in the first year, when two observations are required but not prior to the file deadlines).
- One of your observations must be of a **General Studies course**.
- Our official procedures state that “Up to two additional observations may be requested by the probationary faculty member”—but remember that you can have as many formative observations as you like.



# Schedule observations early

- I recommend **one formative observation in the first semester for all new faculty**, whether part-time, full-time visiting, or full-time tenure track.
- In the spring, I recommend that observations be scheduled fairly early in the term to allow time for a redo or rescheduling.

# Choose your observer

Policy: Observers must be **tenured faculty members** chosen by consensus of faculty member to be observed, Dean, and a tenured faculty member mutually selected by the faculty member and the Dean.

## Recommendations:

- Have at least one observation of a program course by a member of your program
- Have at least one observation by someone from outside your program
- If you were hired particularly for a course or set of courses, schedule at least one observation of at least one of those courses
- Have the Director of the Institute for Faculty Development complete at least one observation before your tenure file is due
- Choose observers who you trust to complete your observation write up fairly and in a reasonable amount of time.
- Choose observers who you respect—whose advice about your teaching you will listen to

# Prepare to be observed (info on IFD website)

- Ask observer at least a week **in advance** of the class to be observed.
- Let the observer know if your program has special procedures or forms.
- **Provide a copy of a syllabus, assignments, and other course materials** (policy).
- Provide a copy of your program standards and **remind the observer to connect their observations to the college, school, and program standards.**
- Provide **context for the course** (number of times you've taught, major revisions made, changes in program, etc.)
- Provide **context for the class** (what done before, what leading up to, etc.)
- If possible, **meet with observer before the observation** to discuss these contexts.
- Let the observer know any things you'd particularly like them to attend to and any other requests—where you'd like them to sit, whether it is ok if your observer brings a beverage, if it is ok for them to type or not, etc.
- **Prepare, of course, but do the best of your typical teaching.** I don't recommend trying something you don't usually do—you may not be able to foresee problems, your students may react oddly....
- If something unusual happens, let the observer know right away and arrange a time to discuss it further or have a second observation

# **IMPROVING TEACHING AND EVALUATIONS**

# Use the data on your student evaluations and from your observations.

- Teachers can look to the information on page three of IDEA reports to see what steps they might take to improve student progress on various objectives.
- Teachers can talk with peer observers.
- Research strongly indicates that teachers who consult with someone about results are more likely to see improvement in their results in the future.

# Improving Evaluations

- **Students will more likely perceive progress if you relate to them, early and often, what goals they are working on and use vocabulary similar to that on the student evaluation forms.** This means that your syllabus, assignments, and in-class discussion/lecture should refer to the goals and help students see how activities are intended to help them make progress on the goals.

# Improving Evaluations

- **Use page four on your student evaluation forms to track student feedback over time**—the items in bold on page 4 of IDEA or mentioned by students on the small class form should match the items on which you're receiving the highest scores if your students perceive that they have made progress on the learning objectives you selected for the class. High scores or frequent occurrence of items you didn't select may point to things you might add later. Low scores or lack of appearance of items you did select point to a need to make a change, either in your choice of objectives, your course, your pedagogy, or a mix of these.

# TEACHING PORTFOLIO



# What it should include (years 2 and on)

- ❑ Your formal student evaluations
- ❑ Your preceptorial advising evaluations
- ❑ Your teaching observation write-ups
- ❑ At least one syllabus from a program course
- ❑ At least one syllabus from a G course
- ❑ Several sample assignments or other course materials
- ❑ If you like, midterm evaluations
- ❑ If you have them, other assessments of student learning
- ❑ A description of your teaching philosophy, in which you connect what your students and peers say, your syllabi, and some sample in-class or homework assignments to the your philosophy and to the college, school, and program standards. In this description, you should draw upon qualitative and quantitative data to **show**, not tell, evaluators about your teaching.

# REFERENCES

# References

- Cashin, William. “Student Ratings of Teaching, the Research Revisited.” 1995. Idea paper 32. [http://www.theideacenter.org/sites/default/files/Idea\\_Paper\\_32.pdf](http://www.theideacenter.org/sites/default/files/Idea_Paper_32.pdf)
- Cashin, William. “Student Ratings of Teaching: A Summary of the Research.” 1988. Idea paper 20. [http://www.theideacenter.org/sites/default/files/Idea\\_Paper\\_20.pdf](http://www.theideacenter.org/sites/default/files/Idea_Paper_20.pdf)
- Colman, Andrew, Norris, Claire., and Preston, Carolyn. “Comparing Rating Scales of Different Lengths: Equivalence of Scores from 5-Point and 7-Point Scales.” 1997. Psychological Reports 80: 355-362.
- Hoyt, Donald and Pallett, William. “Appraising Teaching Effectiveness: Beyond Student Ratings.” Idea paper 36. [http://www.theideacenter.org/sites/default/files/Idea\\_Paper\\_36.pdf](http://www.theideacenter.org/sites/default/files/Idea_Paper_36.pdf)
- “Interpreting Adjusted Ratings of Outcomes.” 2002, updated 2008. <http://www.theideacenter.org/sites/default/files/InterpretingAdjustedScores.pdf>
- Pallet, Bill. “IDEA Student Ratings of Instruction.” Stockton College, May 2006.
- “Using IDEA Results for Administrative Decision-making.” 2005. <http://www.theideacenter.org/sites/default/files/Administrative%20DecisionMaking.pdf>