The Assessment of Teaching in Higher Education

Philip C. Abrami

Centre for the Study of Learning and Performance

Concordia University, Montreal, Canada

Three key questions

- Is it useful and important to evaluate teaching?
- Are imperfect teacher evaluations better than no evaluations at all?
- Can research evidence help improve the evaluation process?

The Teaching Dossier

- Course outlines
- Presentation notes
- Evidence of student learning (e.g., sample term papers, exam results, etc.)
- Chair or senior faculty classroom visits
- Alumni or graduate feedback
- Microteaching vignettes
- Faculty development efforts
- **Student ratings of teaching effectiveness (aka Teacher Rating Forms [TRFs])**

- How is this evidence used to judge the quality of teaching?

Reasons for using TRFs

- Time and cost efficiency
- Flexibility and comparability across multiple teaching contexts
- Students as consumers
- Reliability
  - Test-retest and internal consistency in the range of .60-.80 when there are 15 student raters or more.
- Validity

Concerns about TRFs

- Teaching is multidimensional so TRFs must be multidimensional

- Ratings cannot be used to measure an instructor's impact on student learning

- Ratings are popularity contests

- Instructors who assign high grades are rewarded by high ratings

- Other extraneous factors introduce bias into ratings that make interpretation unfair.

- Review committees do not evaluate teaching properly

- TRFs cannot be used for teaching improvement

- Electronic teaching portfolios are a waste of time.

Select Faculty Reactions to Student Ratings

TRFs: Global questions (Product)

- How would you rate this instructor in general, all-around teaching ability?

- How much did you learn in this course?

- How would you rate the overall effectiveness of this course?

TRFs: Specific questions (Process)

- Understandable course objectives
- Communicates clearly
- Uses appropriate evaluation techniques
- Gives adequate feedback
- Is well prepared
- Is enthusiastic
- Answers questions
- Permits differing points of view
- Is accessible
- Makes it easy to get help

Global vs. specific questions

- Global questions very useful for summative decisions
- Specific questions low in content and predictive validity across myriad teaching situations
    - Large lecture vs. small discussion
    - Teacher vs. student centred learning (collaborative learning, problem-based inquiry)
    - Distance, online and blended learning
- Specific questions very useful for formative purposes

Multisection validation studies

- Multiple sections of the same course; different instructors
- Common teaching materials
- Common examinations (usually multiple choice tests)
- Random assignment of students or pretest equivalence
- What is the relationship between mean TRF scores and teacher-produced student learning

43 Multisection Validity Studies

- Multivariate meta-analysis (d'Apollonia & Abrami;1996, 1997a, & 1997b)
  - 741 validity coefficients
  - General instructor skill = +.26 to +.40 (95% CI)
  - Correction for attenuation = +.47
- Cohen (1981)
  - Specific factors
  - Validity coefficients were lower (e.g., -.02; +.23)

  - **Global rating items are better predictors of teacher-produced student learning than specific rating items**

Bias in student ratings

- Irrelevant influences on student ratings
- More specifically, influences on student ratings that are different from the influences on teacher-produced student learning.

- Examples: **teacher personality**, **grading policies**, elective vs. non-elective course, course level, student major vs. non major, teacher and student gender, class size, and so on.

The Doctor Fox Effect

- Educational Seduction

- What is the effect of teacher personality, more specifically, instructor expressiveness, on student ratings?

- Is their a "biasing" effect of instructor expressiveness even when lecture content is low?

The Original Dr. Fox Effect

Experimental Research on the Dr. Fox Effect

- Instructor Expressiveness (High, Low)

- Lecture Content (High, Low)

- Measures (Student ratings, Learner achievement)

- Meta-analysis results of 12 studies (Abrami et al., 1982)

	TRF	ACH
Expressiveness	.285	.043
Content	.046	.158

- **Results help explain the moderate correlation between student ratings and achievement.**

Influence of Grades on Student Ratings

Do Instructor Grading Standards influence TRFs?

- There is a moderate correlation between mean course grade and mean instructor TRFs (Greenwald & Gilmore, 1997)

- But is this relationship attributable to differences in teacher-produced student learning or to differences in teacher grading policies?

- The learning hypothesis says that differences in ratings are attributable to teacher impacts on student learning and are valid influences.

- The grading bias hypotheses says that differences in ratings are attributable to teacher grading practices and are invalid influences.

- In the only experiment ever conducted on the influence of grading practices, Abrami et al. (1980) found that teachers low in both expressive and lecture content, received **worse ratings** when they assigned **higher grades**.

Summative Decisions about Teaching Using TRFs

- TRFs are psychometrically defensible measures of teacher effectiveness.

- But ratings are only moderate, imperfect predictors.

- Therefore, the *validity coefficient* for ratings should be taken into account when accumulating and interpreting evidence regarding summative decisions about teaching effectiveness.

- *Normative decisions* about teaching (i.e, faculty compared to other faculty) should be based on evidence accumulated over multiple course and formed into a statistical *confidence interval*.

- *Criterion decisions* about teaching (i.e., faculty judged against a standard) should also be based on accumulated evidence and formed into a *confidence interval*, not only a point estimation.

Norm-based procedures adjusted by VC

$$t_{v\dot{c}} = \frac{\bar{Y}_i - \bar{Y}_g}{\sqrt{\left(\dfrac{s_i^2}{n_i} + \dfrac{s_g^2}{n_g}\right)\left(\dfrac{1}{1 - vc}\right)}} \quad \text{for} \quad df = n_i + n_g - 2$$

where $\bar{Y}$ is the mean TRF score, s^2 is the unbiased variance, n is sample size, vc is the validity coefficient, and df is the degrees of freedom.

In addition, one can calculate a confidence interval for the calculated value of t_{vc}:

$$CI = (\bar{Y}_i - \bar{Y}_g) \pm t_\alpha s_{Dvc}$$

where t_α is the critical value of t at a particular alpha level and

$$s_{Dvc} = \sqrt{\left(\frac{s_i^2}{n_i} + \frac{s_g^2}{n_g}\right)\left(\frac{1}{1 - vc}\right)}.$$

Recommendations for Summative Decisions using TRFs

1. Report the average of several global items.
2. Combine the results of each faculty member's courses.
3. Decide in advance on the policy for excluding TRFs (e.g., new courses, classes with small ns, etc.).
4. Choose between norm-referenced and criterion-referenced evaluation and the level of acceptable performance.
5. Follow the steps in statistical hypothesis testing.
6. Provide descriptive and inferential statistics and illustrate them in a visual display.
7. Incorporate TRF validity estimates into statistical tests and confidence intervals.
8. Use class mean TRFs not individual students as the units of analysis.
9. Decide whether and to what extent to weigh other sources of evidence.

Table of specifications rating scale

Use the following rating scale in making your judgments:
1—Poor
2—Fair
3—Good
4—Very good
5—Excellent
NA—*Not applicable*

Course Content	How the Content Was Learned						OVERALL RATING
	Knowledge	Compre-hension	Applica-tion	Analysis	Synthesis	Evaluation	
Descrip-tive statis-tics							
The t-test							
Oneway Anova							
Factorial Anova							
Nonpara-metrics							
OVER-ALL RAT-ING							

Recommendations for Formative Decisions about TRFs

- Cohen's (1980) meta-analysis showed that feedback from student ratings in one course improved ratings in a subsequent course.
- For best results:
  - Specific aspects of teaching should be explored via a cafeteria system.
  - If ratings are collected part-way during a course, results should be discussed with students.
  - Using a faculty development officer as a consultant helps make student feedback useful and interpretable.

Electronic portfolios: Modern tools for faculty assessment

- Not only for students
- Multimedia containers
- Showcase portfolios (summative assessment)
- Process portfolios (formative assessment)
- Self-regulated learning (and teaching)

- Individuals high in SRL outperform those low in SRL on demanding, complex, and novel tasks (Zimmerman, 2011).

- Using ePEARL: Electronic Portfolio Encouraging Active Reflective Learning

The SRL Model

Schunk & Zimmerman (1994, 1998)

SRL Stages

- Forethought or *Planning*
  - Goal setting and selecting task strategies
    - (Goal—increase in-class collaboration)
    - (Strategy—use TGT method)
- Performance/Volitional Control or *Doing*
  - Enacting the task and strategies
    - (Managing the strategy in class)
- Self-reflection or Reflection
  - Self and other feedback
    - (Were students engaged? Did they learn the content? How do I gather student feedback? How do I improve the exercise in future?)

ePEARL Level 4: Using technology for teaching

Thank you

Philip C. Abrami

abrami@education.concordia.ca

http://www.concordia.ca/research/learning-performance.html

The SRL Model

Task Analysis
setting goals
planning strategies

Self-Motivation
self-efficacy
outcome expectations
intrinsic interest
learning goal orientation

Performance or
Volitional Control

Processes which occur during learning
efforts and affect attention and action

Forethought

Influential processes which
precede efforts to act and
set the stage for action

Schunk & Zimmerman (1994, 1998)

The SRL Model

Performance or Volitional Control — Processes which occur during motoric efforts and affect attention and action

Self-Control
self-instruction
imagery
attention focusing
task strategies

Self-Observation
self-recording
self-experimentation

Schunk & Zimmerman (1994, 1998)

The SRL Model

Self-Judgment
  self-evaluation
  causal attribution

Performance or
Volitional Control

Processes which occur during motoric
efforts and affect attention and action

Self-Reaction
  self-satisfaction/affect
  adaptive-defensive responses

Forethought

Influential processes which
set the stage for action

Self-Reflection

Processes which occur after
performance efforts and influence a
person's response to that experience
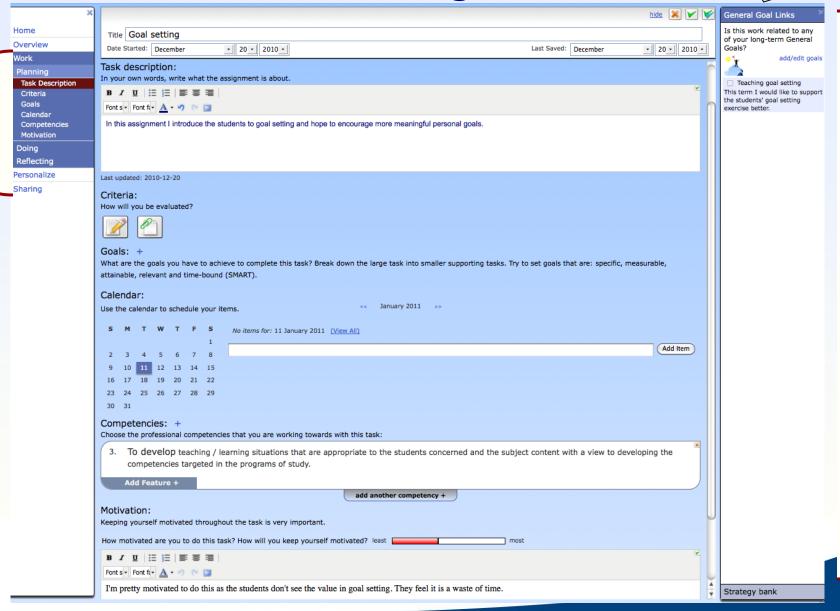
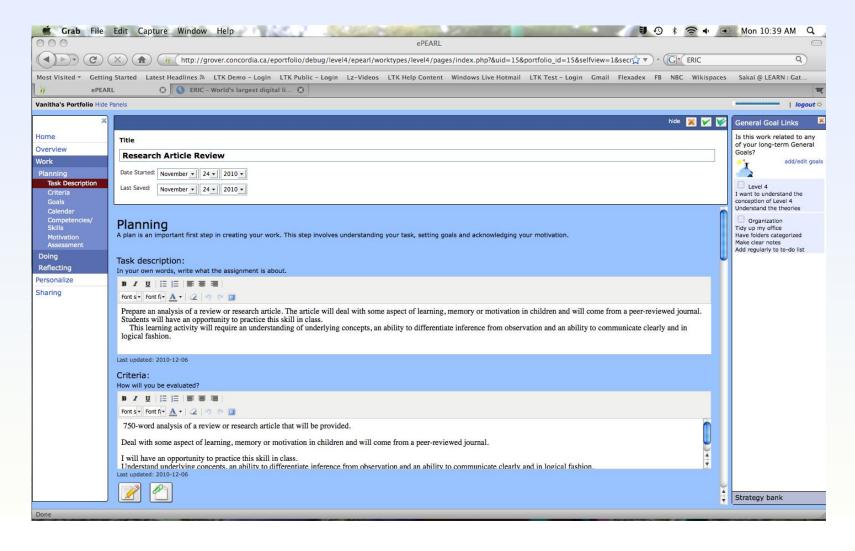Schunk & Zimmerman (1994, 1998)

More on ePEARL

ePEARL Level 4: Planning

Planning – Task Description & Criteria

Task Goals & Supporting tasks

Task Analysis
    setting goals
    planning strategies

Visualizing – Concept Map

Strategy Bank

Calendar

Calendar:

Use the calendar to schedule your items.

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26

No items for: 02 February 2011   (View All)

Mozilla Firefox

http://grover.concordia.ca/eportfolio/debug/level4/epearl/worktypes/level4/pages/calendar.php?mode=single&entry_id

Date	Item	Completed
	Set a date fro Training	☐
	Review with Einat	☐
	Potential training date	☐
	Meet w/ Anne and go over	☐
	Practice	☐
	Presentation Date	☐

Calendar

<<   **February 2011**   >>

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28					

Items for: 02 February 2011

Teaching Resource – Summarizing Revise
Teaching Resource – Summarizing v2 Revise

Upcoming:

02/02/2011 **Teaching Resource – Summarizing**
Revise
02/02/2011 **Teaching Resource – Summarizing**
v2

Competencies

Title: ePEARL Level 4 - Workshop

Date Started: February | 2 | 2011   Last Saved: February | 2 | 2011

Competencies: +

Choose the professional competencies that you are working towards with this task:

2. To communicate clearly in the language of instruction, both orally and in writing, using correct grammar, in various contexts related to teaching.

    - Uses appropriate language when speaking to students, parents and peers.

    - Is able to take up a position, support his or her ideas and argue his or her subject matter in a consistent, effective, constructive and respectful way during discussions.

    Add Feature +

11. To engage in professional development individually and with others.

    Select a feature:
    - Takes stock of his or her competencies and takes steps to develop them using available resources.
    - Discusses the relevance of his or her pedagogical choices with his or her colleagues.
    - Reflects on his or her practice (reflective analysis) and makes the appropriate adjustments.
    - Spearheads projects to solve teaching problems.
    - Involves peers in research related to the mastery of the competencies targeted in the programs of study and to the educational objectives of the school.

Select a competency:

1. To act as a professional who is inheritor, critic and interpreter of knowledge or culture when teaching students.
3. To develop teaching / learning situations that are appropriate to the students concerned and the subject content with a view to developing the competencies targeted in the programs of study.
4. To pilot teaching/learning situations that are appropriate to the students concerned and to the subject content with a view to developing the competencies targeted in the programs of study.
5. To evaluate student progress in learning the subject content and mastering the related competencies.
6. To plan, organize and supervise a class in such a way as to promote students' learning and social development.
7. To adapt his or her teaching to the needs and characteristics of students with learning disabilities, social maladjustments or handicaps.
8. To integrate information and communications technologies (ICT) in the preparation and delivery of teaching/learning activities and for instructional management and professional development purposes.
9. To cooperate with school staff, parents, and partners in the community and students in pursuing the educational objectives of the school.
10. To cooperate with members of the teaching team in carrying out tasks involving the development and evaluation of the competencies targeted in the programs of study, taking into account the students concerned.
12. To demonstrate ethical and responsible professional behaviour in the performance of his or her duties.

Motivation

Level 4: Doing

Sidebar

Sidebar

Self-Observation
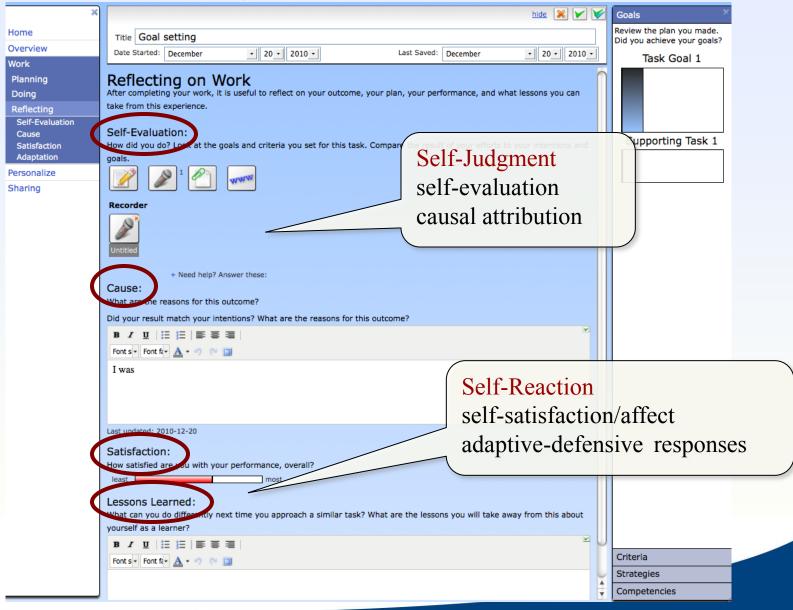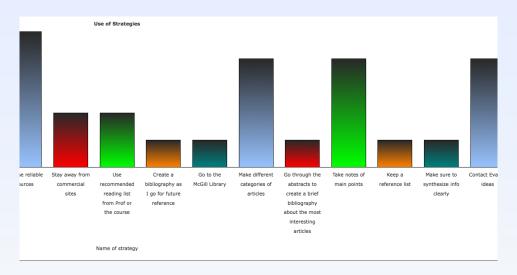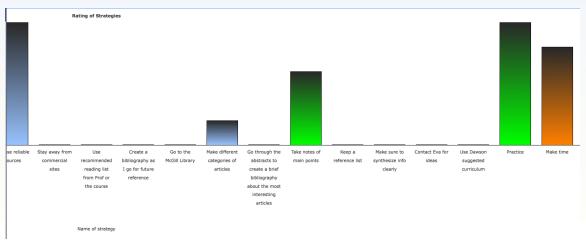self-recording
self-experimentation

Teacher Sharing

Level 4: Reflecting

Overview - Graphs

Overview – Strategy Management