

The Discipline of Evaluation: Teaching Expert Reading

Eleanor Wittrup
Department of Philosophy

The moment I got interested in assessment...

DF: Do you and your colleagues have a shared standard for grading students?

Me: Of course!

DF: Wouldn't it be fair to tell the students what it is?

Me: Er....

Philosophy = Extreme Honesty (plus Respect, Compassion, Fairness, Responsibility and a dollop of Courage)

7 Elements of All Evaluation

- Michael Scriven

- Purpose of Evaluation (What do you want it for?)
 - Improve a practice (formative)
 - Make a judgment of merit/worth/significance (summative)
 - Improve our understanding of something (ascriptive)
- Context of Evaluation:
 - Time horizon
 - Circle of concern
 - Environment
- Evaluand
- Practical Method of Evaluation
- Criterion or Criteria (and for each criterion we need a)
- Standard(s) of evaluation:
 - Ranking/scoring
 - Comparison group/scale
- Results of Evaluation
- Conclusion (depending on type of evaluation) Michael Scriven

The Problem: Students read poorly

- They don't understand the position
- They don't understand the argument
- They get confused by counterarguments
- When they get lost they give up
- ***They don't even try to genuinely understand what they are reading***

Purpose of Evaluation:

Improve student reading (formative)

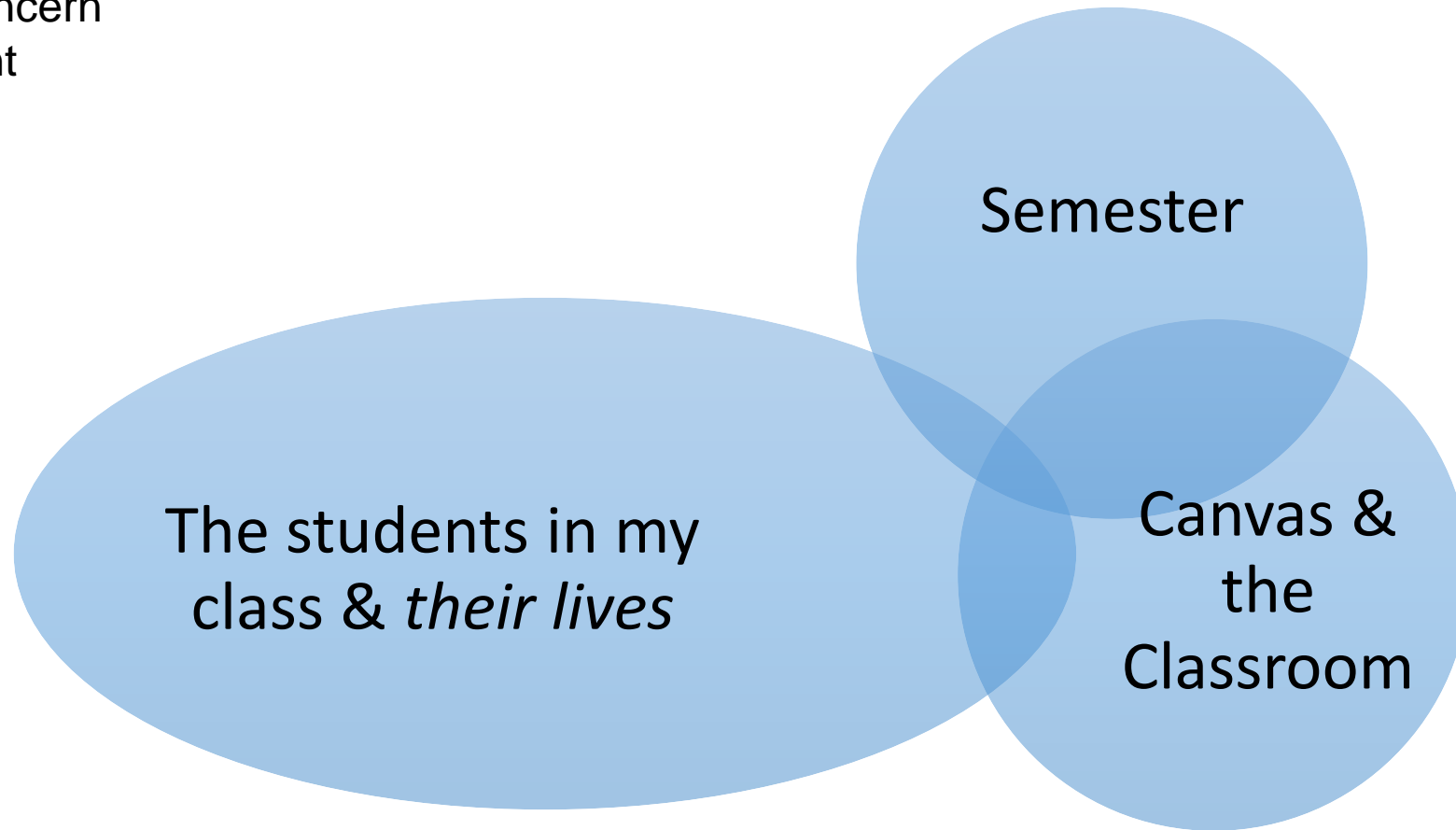
- Students care about understanding the reading/getting it right
- Students get a sense of the author's intent
- Students recognize what they do and don't "get" in a reading
- Students make connections between the content of the reading and their lives/interests
- **AUTHENTIC ASSESSMENT:** Give them a complicated essay "cold" and see if they can understand it by themselves.

Context of evaluation: My class

Time horizon

Circle of concern

Environment



Using Prep Sheet for Evaluation

- Evaluand: Prep Sheet
- Purpose of Evaluation: (formative) Improve student reading
- Context of Evaluation: 1 day, this student & reading, this class
- Practical Method of Evaluation: read prep sheet and fill our rubric
- Criteria & Standard(s) of evaluation: (scoring) see Canvas rubric
- Result: Scores on prep sheets improve over time
- Conclusion: Student reading improves

The Evaluand: The prep sheet

(NOT the student)

- An impersonal method of communicating with the students about their performance on the skill I wish them to acquire.
- Leave personality OUT of it.

Practical Method:



- Require a prep sheet for each class reading
- Read each prep sheet on Canvas
- SCORE/Fill out rubric for each prep sheet
- Require students to discuss reading knowledgably in class with each other
- Offer a “higher level” prep sheet option

Criteria: What I evaluate





- Ability to identify examples (concrete points of contact with the world of shared experience)
- Ability to explain what the examples are used for
- Ability to identify key distinctions and explain them
- Ability to define the key terms and context of the discussion
- Ability to identify and clearly articulate the author's main claim/focus
- Ability to identify and describe the structure of the argument
- Ability to ask sophisticated critical questions about an essay

Type of Evaluation: Scoring yielding a profile

<p>1. Example Identification threshold: 3.0 pts</p>	<p>Only the most important examples correctly listed 4.0 pts</p>	<p>All the examples correctly listed 3.0 pts</p>	<p>All the examples are examples and most of the important ones are included 2.0 pts</p>	<p>Some members of the list are not examples 1.0 pts</p>	<p>No list OR No examples on the list 0.0 pts</p>
<p>2. Example Explanation threshold: 3.0 pts</p>	<p>All examples are insightfully explained 4.0 pts</p>	<p>All examples are explained correctly 3.0 pts</p>	<p>All examples are explained, only some are correct. 2.0 pts</p>	<p>Some examples are explained, others not. OR generic answers are given 1.0 pts</p>	<p>No explanations OR meaningless explanations 0.0 pts</p>

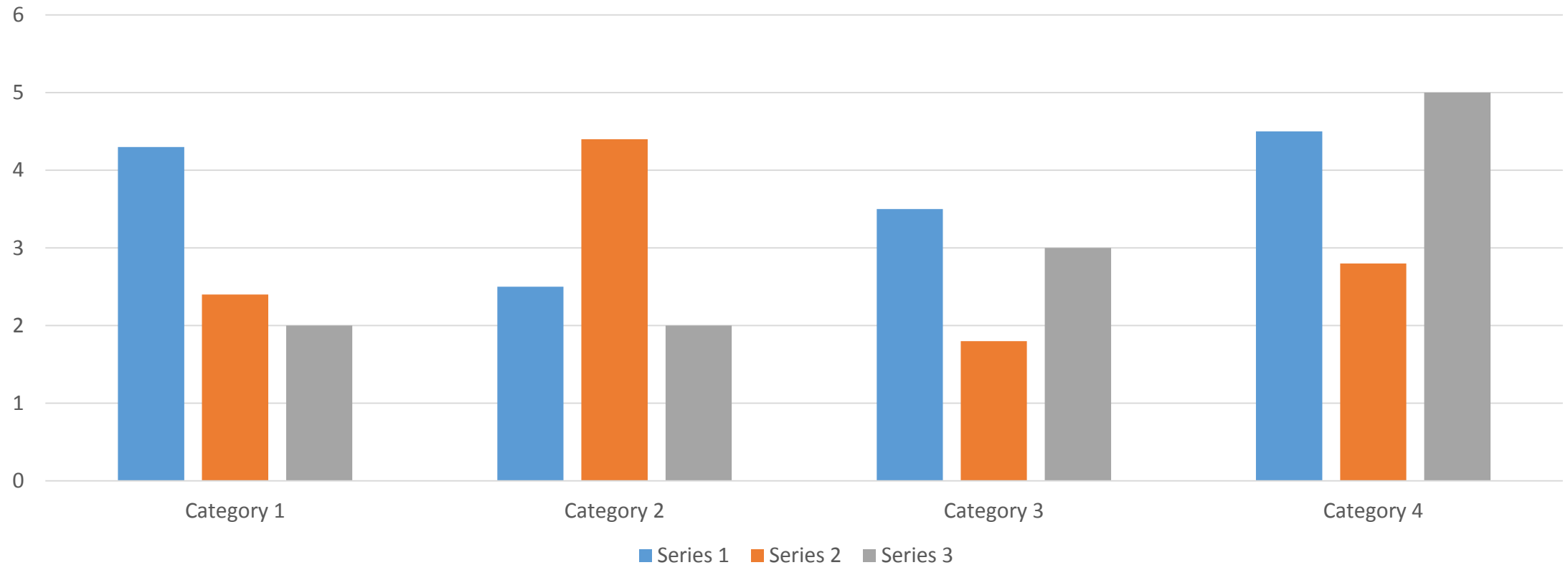
Criteria	Ratings										Pts
<p> Example Identification & Explanation view longer description threshold: 4 pts</p>	<p>Use of examples furthers the argument and demonstrates mastery of the topic. 10 pts</p>	<p>Most important examples are listed, explained, and their strength evaluated. 9 pts</p>	<p>Most important examples are listed, explained and how they create the context of the discussion is made clear. 7 pts</p>	<p>Most important examples listed and their point for the essay explained. 6 pts</p>	<p>Most important examples listed and identified as evidence or illustration 5 pts</p>	<p>Only the most important examples listed 4 pts</p>	<p>All the examples correctly listed 3 pts</p>	<p>All the examples are examples and most of the important ones are included 2 pts</p>	<p>Some members of the list are not examples 1 pts</p>	<p>No list OR No examples on the list 0 pts</p>	<p>10 pts</p>
<p> Making Distinctions view longer description threshold: 4 pts</p>	<p>Makes compelling new distinctions or used existing distinctions in a new or surprising way. 10 pts</p>	<p>Correctly identifies, explains and contextualizes the most important distinctions and evaluates them. 8 pts</p>	<p>Correctly identifies and explains the most important distinctions, and puts them into the context of the argument or explains how they contextualize the argument. 6 pts</p>	<p>Correctly identifies and explains the most important distinctions 5 pts</p>	<p>Correctly identifies some of the most important distinctions and explains them but can't explain them all accurately. All the distinctions identified are explained. 4 pts</p>	<p>Correctly identifies distinctions (but may miss the most important ones) and explains them all 3 pts</p>	<p>Correctly identifies some of the most important distinctions but cannot explain them. 2 pts</p>	<p>Identifies but doesn't explain some distinctions, or doesn't identify the most important ones. 1 pts</p>	<p>Cannot correctly identify distinctions being made 0 pts</p>	<p>10 pts</p>	

 Structuring Arguments view longer description threshold: 3 pts	Structures argument clearly, coherently and effectively. 10 pts	7 points plus evaluates soundness of argument. 9 pts	5 points plus evaluates validity of argument. 7 pts	Sections appropriate, titles help follow the argument, structure explained and contextualizes argument 5 pts	Appropriately breaks down reading into sections and names accurately, explains how this structure helps you follow the argument. 4 pts	Appropriately breaks down reading into sections and names accurately 3 pts	Sections appropriate, titles confusing, inaccurate or repeat existing titles 2 pts	Section breaks confusing/arbitrary AND/OR No titles 1 pts	No sections OR restatement of existing ones 0 pts
 Making Claims view longer description threshold: 5 pts	Main claim identified and clearly stated, and key terms identified. All key terms defined in a way that contextualizes the discussion. The claim made by the student is controversial and is a contribution to the wider discussion. 10 pts	Main claim identified and clearly stated, and key terms identified. All key terms defined in a way that contextualizes the discussion. The claim made by the student is controversial and demonstrates broad knowledge of the topic. 9 pts	Main claim identified and clearly stated, and key terms identified. All key terms defined in a way that contextualizes the discussion. 7 pts	Main claim identified and clearly stated, and key terms identified. Some of key terms defined. 5 pts	Main claim identified and clearly stated, and key terms identified. 4 pts	Main claim correctly identified and clearly stated. 3 pts	Main claim identified but not clearly stated 2 pts	Ambiguous/unclear statement of a claim OR not the main claim 1 pts	No claim identified 0 pts

 Whole Question 1 (Contextualize) view longer description threshold: 3 pts	Exceeds expectations 4 pts	Introduction refers to the reading and correctly reports on it 3 pts	Introduction refers to the reading but does not correctly report it 2 pts		Introduction does not refer to the reading 1 pts	No introduction 0 pts	4 pts
 Whole Question 2 (State question) view longer description threshold: 3 pts	Exceeds Expectations 4 pts	Clear statement of a specific answerable question about the content of the reading 3 pts	Vague statement of an answerable question about the reading 2 pts	Question not directly about the content of the reading 1 pts	Not a question or not important to the content of the reading 0 pts		4 pts
 Whole Question 3 (Imagine possible answers) view longer description threshold: 3 pts	Exceeds Expectations 4 pts	Clear statement of more than two plausible answers. 3 pts	Clear statement of two plausible answers or vague statement of more than two 2 pts	One plausible answer clearly stated or more than one vaguely stated 1 pts	No plausible answers suggested 0 pts		4 pts
 Whole Question 4 (Explain Uncertainty) view longer description threshold: 3 pts	Exceeds Expectations 4 pts	Explanation of uncertainty which makes clear what is necessary to resolve it 3 pts	Explanation of uncertainty which clarifies the issue 2 pts	Explanation of the uncertainty that is vague or difficult to understand 1 pts		No explanation 0 pts	4 pts

Profile: sliding scale formative evaluation

Student performance on Examples, Distinctions, Claims



Results: Prep sheet scores rise then level out (*but we don't actually care about THAT.*)

- Students complain. A lot initially. It is a huge amount of work at the outset.
- Students drop the class
- Students improve dramatically – reading comprehension improves and then drives class discussion.
- Class discussion and preparation are consistently high
- It takes about four weeks to get into the habit of reading this way, then it is relatively easy.
- Students who become proficient at answering the questions early get bored (“Intermediate” and then “Advanced” prep sheets are then available)
- Students report they begin reading this way in other classes
- Students who finish the class uniformly agree that the prep sheets are essential and extremely valuable and that they read “better.”

Does the Evaluation serve the Purpose?

- Student's reading improves!!! Students:
 - care,
 - don't give up when lost,
 - understand the content (better),
 - can identify the argument,
 - are less confused by counter arguments
 - can talk about the readings intelligently
 - gain confidence in reading

Evaluation as communication:

- To students
 - Is it accurate?
 - Is it informative?
 - Does it help them improve their performance?
 - Does it help them learn?
- To transcript consumers
 - Is it accurate?

The crazy things:

- We needn't give summative synthetic "grades" to individual assignments IF we track individual Outcomes (criteria & standards) and tie final grades to mastery of particular Outcomes and sets of outcomes.
- In fact to do so is *confusing* to students (because it is epistemically disreputable.)
- ***The criteria and standards of our disciplines are not well articulated.***
- Confusion about normal (curved) distributions.

7 Elements of All Evaluation

- Evaluand
- Purpose of Evaluation (For what)
 - Improve a practice (formative)
 - Make a judgment of merit/worth/significance (summative)
 - Improve our understanding of something (ascriptive)
- Context of Evaluation:
 - Time horizon
 - Circle of concern
 - Environment
- Practical Method of Evaluation
- Criterion or Criteria (and for each criterion we need a)
- Standard(s) of evaluation:
 - Ranking/scoring
 - Comparison group/scale
- Results of Evaluation
- Conclusion (depending on type of evaluation)

Bonus Slide:

Three ways of combining several criteria:

- Profile: Reports each criterion of evaluation for an evaluand distinctly in a chart or list.
- Grade: A synthesis of a profile that embodies a judgment of true merit.
- Rank: An ordering of individuals in a defined group.

Bonus Slide: Normal Distribution?

IFF: a randomly distributed attribute in a large AND random AND untreated sample

- Not Normal

- Legs on horses
- Blood type of human beings
- Health status of patients in an ICU
- Ability to ski in people who have taken skiing lessons
- Bacteria count in dogs on antibiotics
- Autism in participants in a conference on advocating for autistic people.
- Level of education of people in a faculty meeting
- Ethnicity in the philosophy dept.
- Grades in a college class

- Normal

- Numbers generated by a large number of roulette wheel spins
- Height of people in Stockton
- Lifetime number of offspring of wild mice
- Snowfall in the Sierra in the last 200 years
- Amount of milk in the fridges of people in Calaveras County