COMPARING MEANS VERSUS PROPORTIONS IN THE ANALYSIS OF END OF COURSE EVALUATION DATA

MURALI KUCHIBHOTLA

AIRUM 2015
NOVEMBER 13, 2015

OUTLINE

- Means and Dispersion Indicator
- Proportions

- Total enrollment of 35,220 students.
- Variety of degrees/programs offered:
  - Bachelor's (26% of total enrollment)
  - Master's (43%)
  - Doctorate/Eds (28%)
  - Certificates (3%)
- Gender
  - 75% female
- Enrollment Status
  - 94% part-time
- Average age of 40 years

MEANS AND DISPERSION INDICATOR

- Nominal Scale: Collections of objects without any sense of order
- Example: Listing of continents (1=Asia, 2=Africa, etc.)
- Ranking these continents by numerical order makes no sense

- Ordinal Scale: Ordered categories
- Ordering makes some comparisons possible such as ranking
- Example: Temperature of a cup of tea
- Natural order: [1]Very Cold [2] Cold [3] Lukewarm [4] Hot [5] Very Hot

- However, other comparisons are not possible with ordinal data
- Addition and subtraction not permitted
- Example: Consider the distance between categories
- Difference 2-1 [cold-very cold] does not equal 3-2 [lukewarm-cold]
- Nor can we say that the average of hot and very hot is **hot and a half!**

- Survey data is ordinal
- Consider this SET statement: “The instructor challenged students to think critically”.
- Students are asked to signal their agreement/dis-agreement with this statement.
- Five point scale: [1] Strongly Disagree [2] Disagree [3] Neutral [4] Agree [5] Strongly Agree
- If 1/2 of a group of n students SA with this statement, while the other half SD, the average is neutral!
- How accurate would it be to claim that students were typically unsure about this statement?

- Means need to be complemented with other measures.
- Such a measure should capture the “dispersion” of responses among students.
- Narrower dispersion indicates more agreement with a statement, compared to wider dispersion.
- Logic of such a measure similar to standard deviation.
- “Consensus” measure, which can be expressed as a percentage, spans the range from 0 to 100.
- Value of 100 indicates perfect agreement amongst students, while 0 indicates complete dissent.

- Hypothetical Scenarios

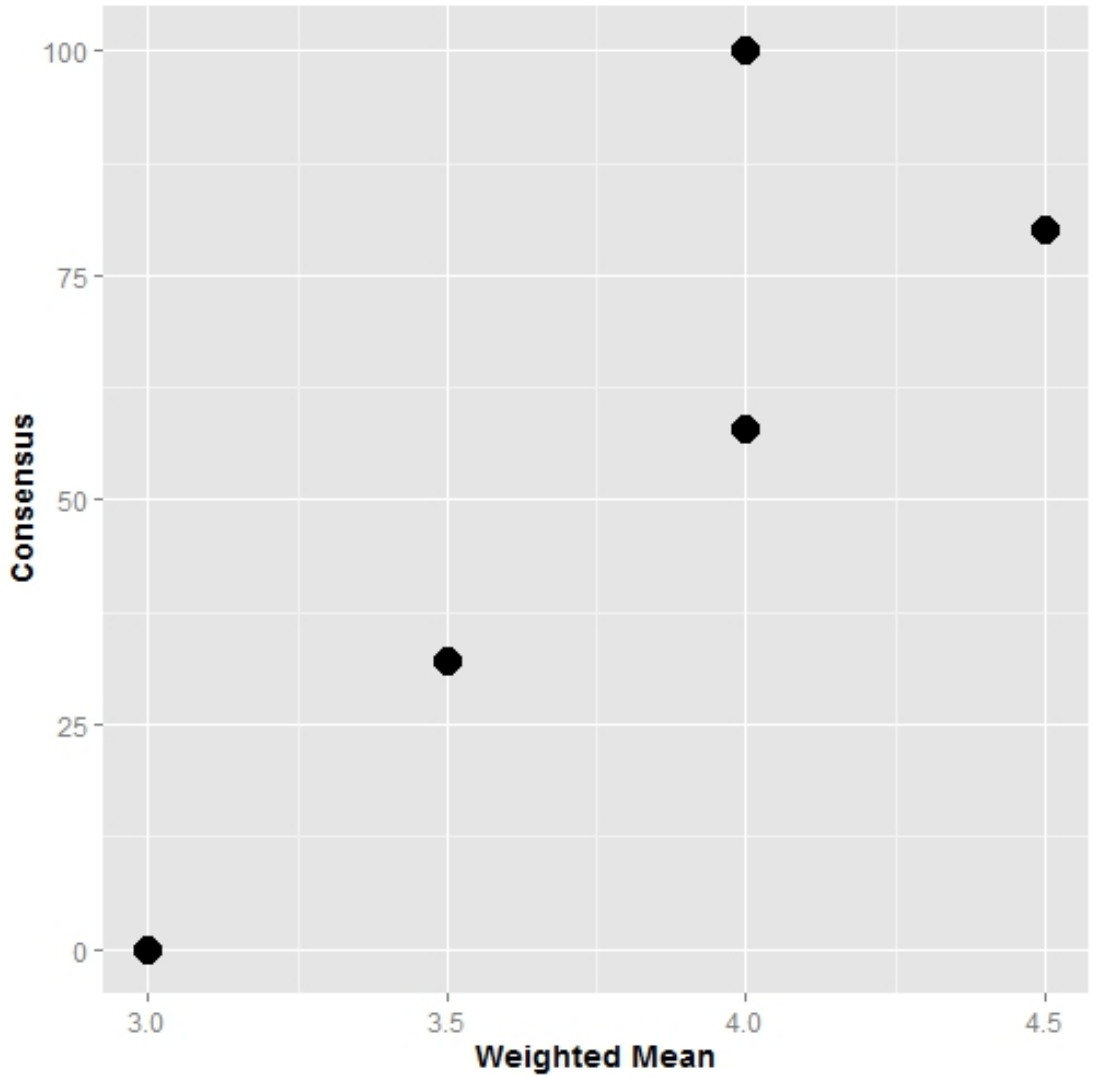
Scenario	SA	A	N	D	SD	Mean
1	6				6	3.0
2		12				4.0
3	6	6				4.5
4	6		6			4.0
5	6			6		3.5

- Measure draws on Claude Shannon's "Fundamental Theory of Information"
- Consensus measure defined as follows:
- $\text{Cns(X)}=1+\sum_{i=1}^{n} p_i \, log_2 \left(1 - \frac{|X_i - \mu_x|}{d_x}\right)$
  - X_i is the particular Likert attribute (i.e., response category).
  - p_i is the probability of the frequency associated with X_i
  - μ_x is the mean of X
  - dx is the width of X.

- Consider scenario 1 from above example
- SD coded as 1, SA as 5.
- X_i =1, 2, 3, 4, and 5
- p_1 =0.5, p_5 =0.5, p_2 =0, p_3 =0, p_4 =0
- μ_x =3
- dx =5-1=4 (difference b/w highest and lowest category)

- Hypothetical Scenarios

Scenario	SA	A	N	D	SD	Mean	Consensus
1	6				6	3.0	0
2		12				4.0	100
3	6	6				4.5	80
4	6		6			4.0	58
5	6			6		3.5	32



PROPORTIONS

- Proportions present a better way of analyzing student evaluation data.
- They have three distinct advantages over means:

1. Valid statistical procedures exist for comparing proportions across different groups
2. Goals set in terms of proportions are easier to understand than those set for means.
3. Using multiple proportion measures allows for greater insights into why instructor performance changes over time.

- Consider a common summary instructor quality question: “How would you rate the quality of the instructor’s online teaching activities?”
- Responses can be : [1] Very Poor [2] Poor [3] Fair [4] Good [5] Very Good [6] Excellent [7] Exceptional
- Suppose an instructor has the following distribution of responses:

N	Very Poor	Poor	Fair	Good	Very Good	Excellent	Exceptional
50	6	5	4	7	8	10	10

- Proportion Measures

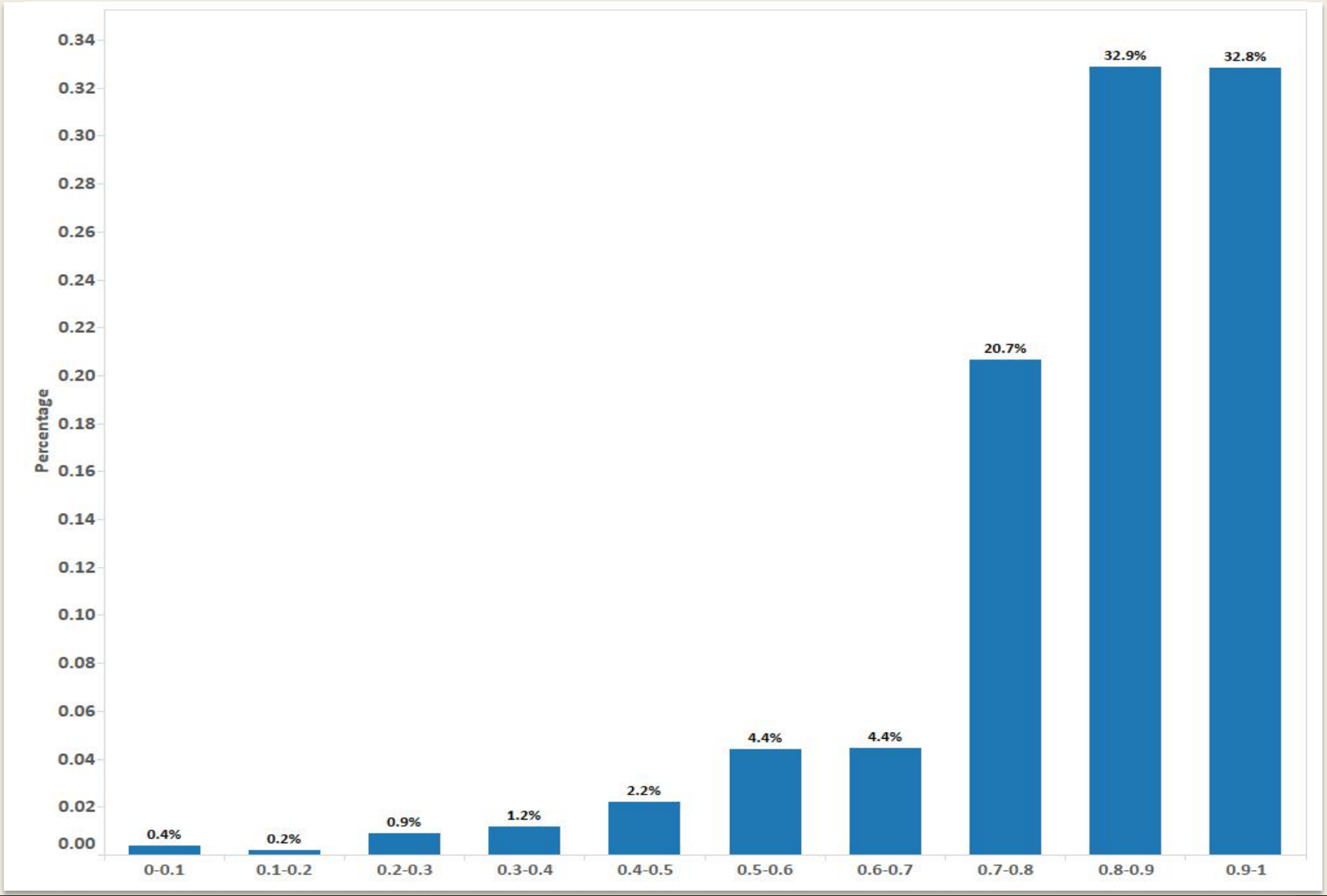
N	P1	P12	P67	P7
50	0.12	0.22	0.40	0.20

- P7 is the proportion of students who answered ‘exceptional’: 10/50=0.20
- P67 is the proportion of students who answered ‘exceptional’ or ‘excellent’: 20/50=0.40
- P1 is the proportion of students who answered ‘very poor’: 6/50=0.12
- P12 is the proportion of students who answered ‘very poor’ or ‘poor’: 11/50=0.22

- Goals set in terms of proportions are meaningful and easily comprehended (by both faculty and administrators).
- Example: “We want our faculty to have 70% of our students answer exceptional or excellent to the summary instructor question.”
- Both faculty and administrators can easily understand what this means.

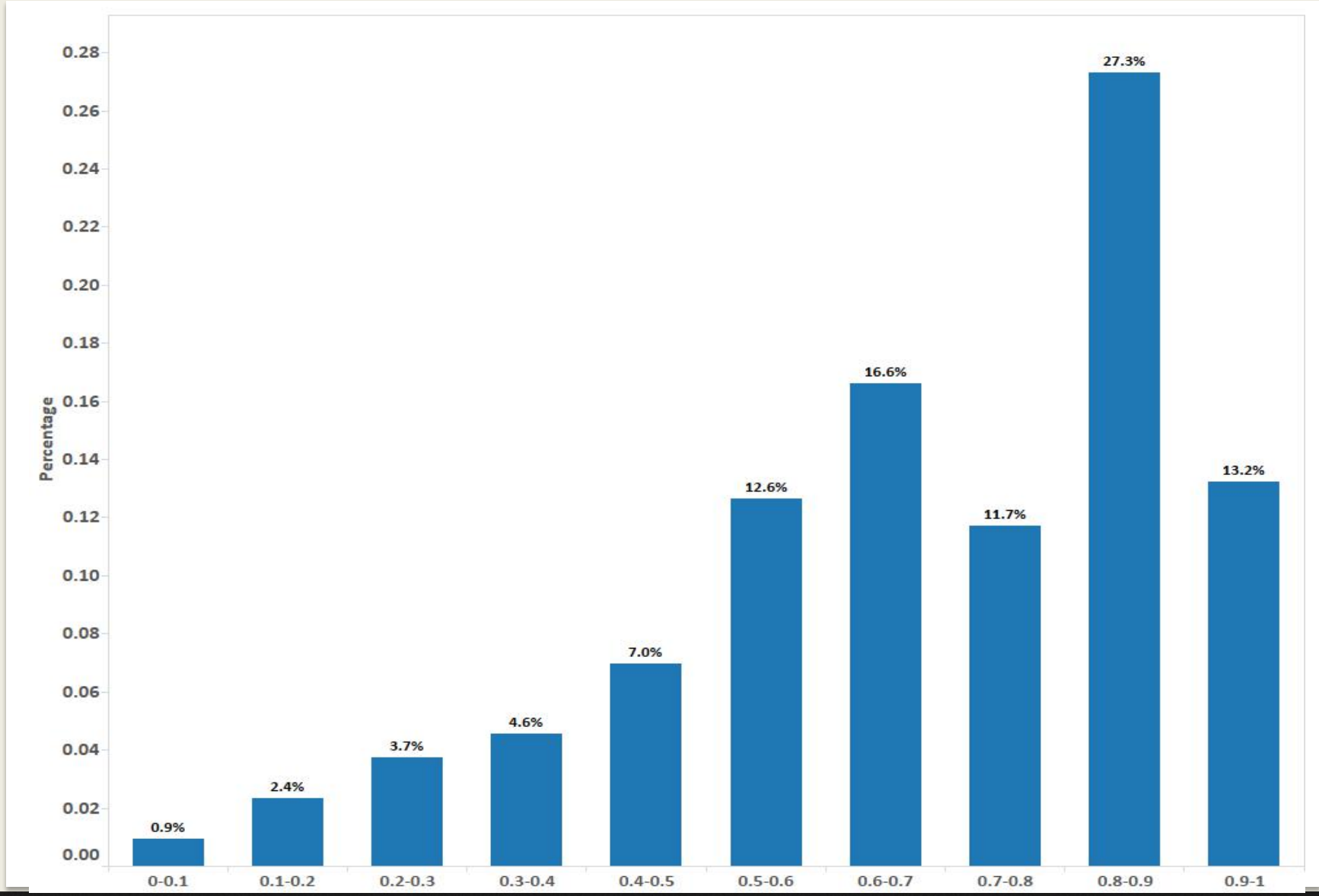
- Which proportion measure should we use?
- Depends on the purpose
- For remedial purposes, p12 useful
- For segmentation purposes (i.e. to identify best faculty) other proportions more useful (such as p67 or p7).
- Statistic should have high variance to serve such a discriminating role.

PROPORTION OF (EXCEPTIONAL, EXCELLENT) RESPONSES



- Evaluation data data are heavily skewed
- 66% of all class sections have *at least* 80% of respondents who rate the instructor as being either exceptional or excellent.
- Most learners quite happy with their instructors
- However, p67 is not a good measure for discriminating amongst classes (or faculty).

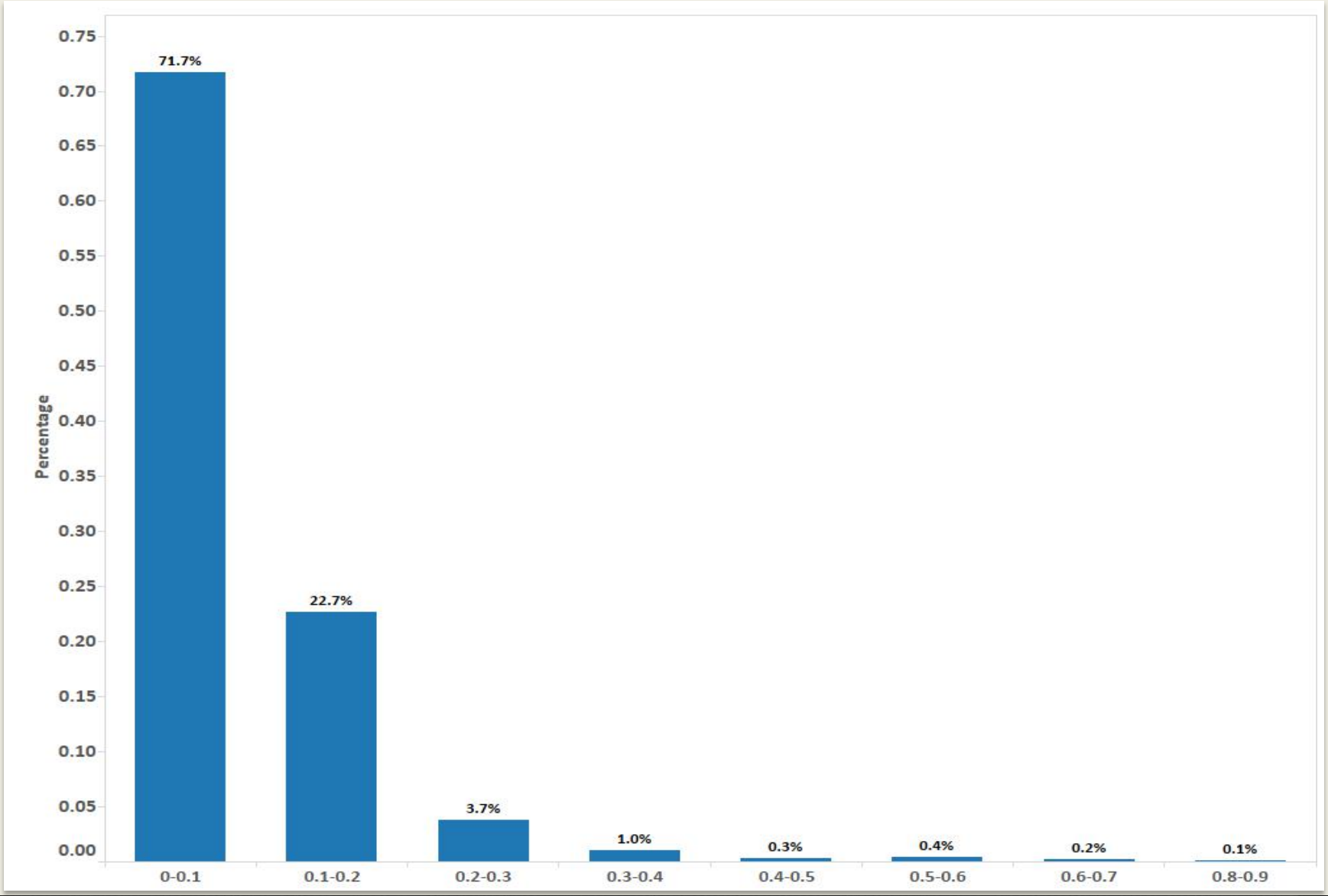
PROPORTION OF EXCEPTIONAL RESPONSES



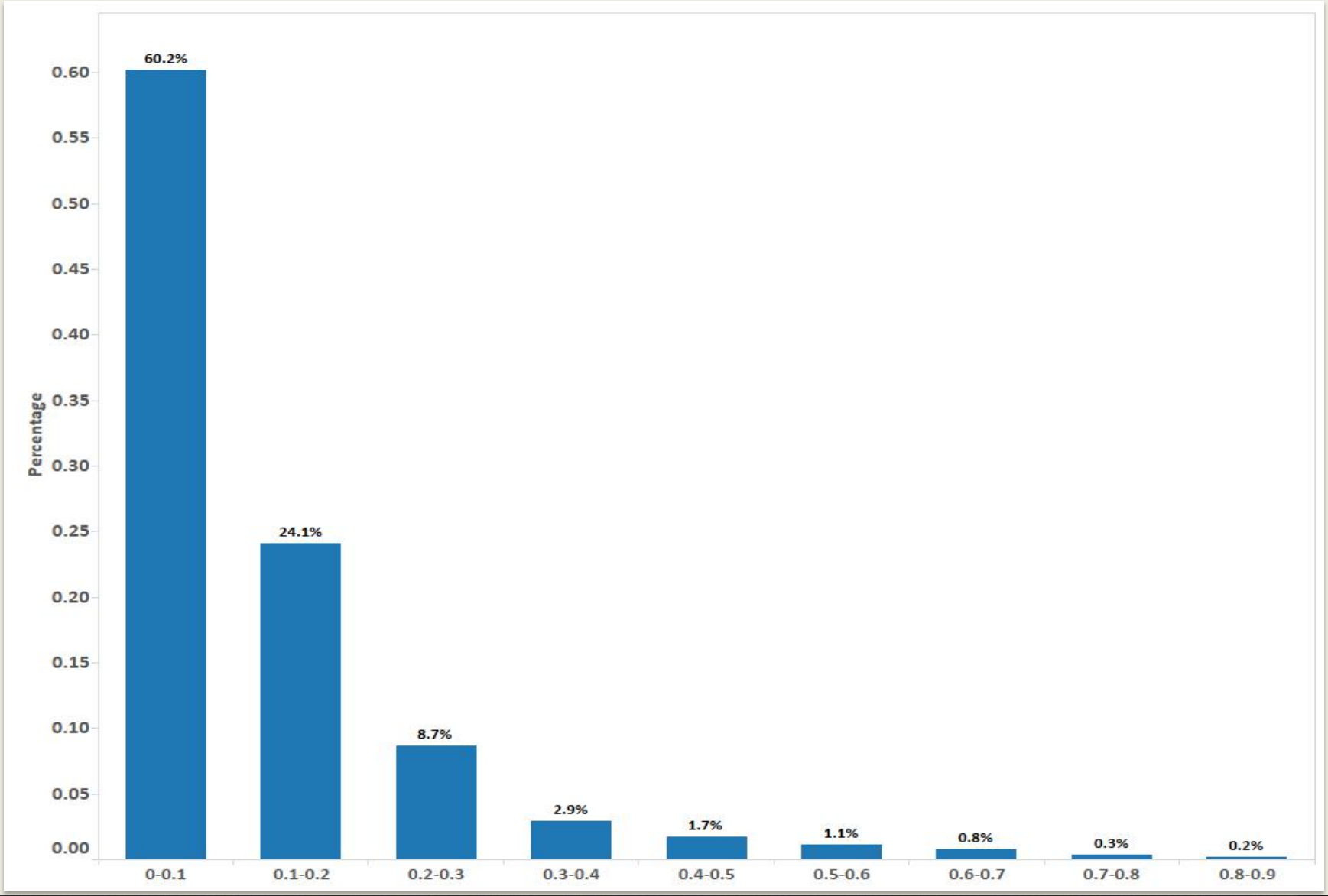
- P7 distribution is far less skewed
- Good choice for discriminating among faculty, to select the best instructors.

- On the other hand, P12 distribution less skewed compared to P1
- Good choice for identifying faculty needing remediation.

PROPORTION OF VERY POOR RESPONSES

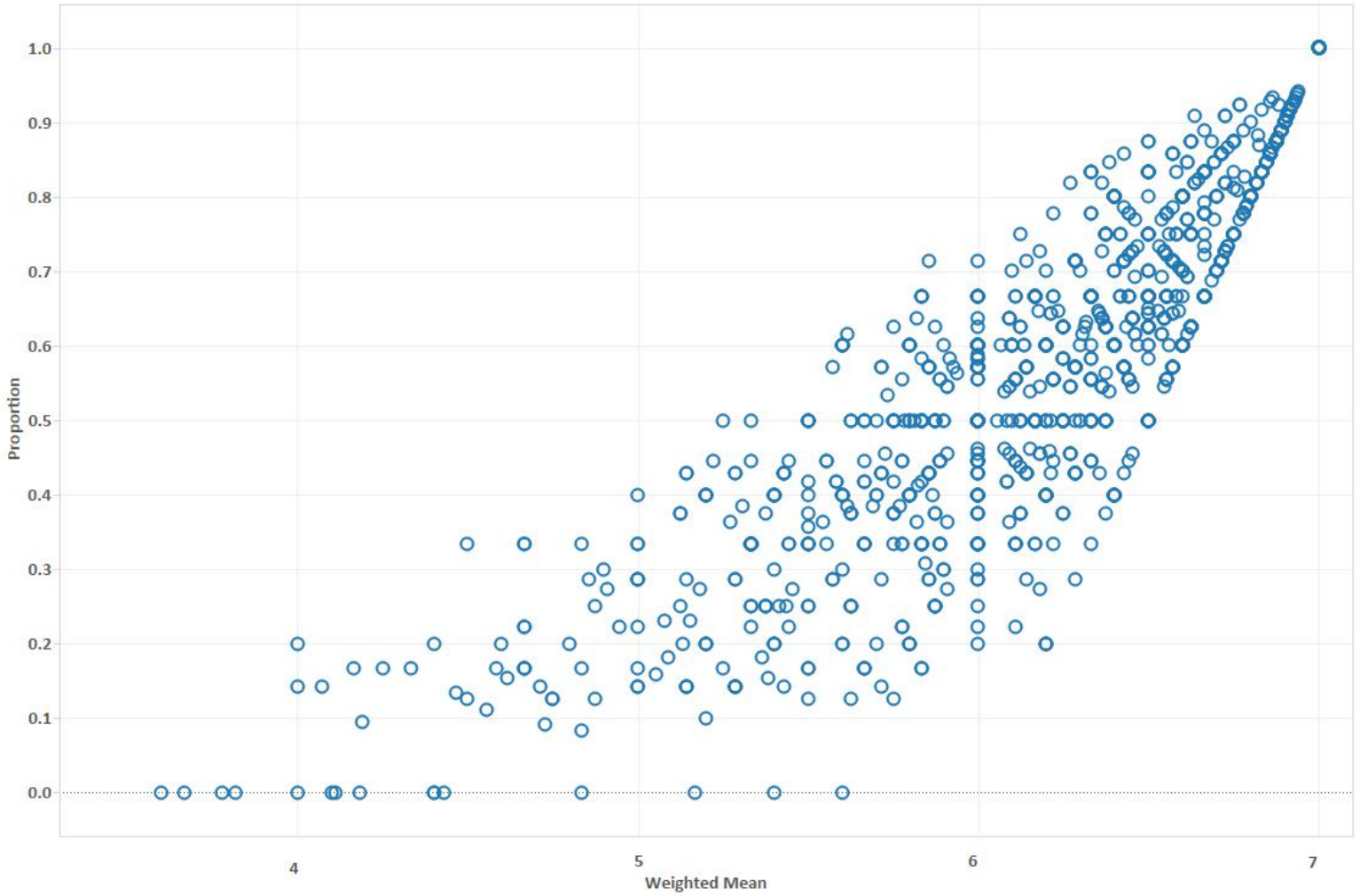


PROPORTION OF (VERY POOR, POOR) RESPONSES

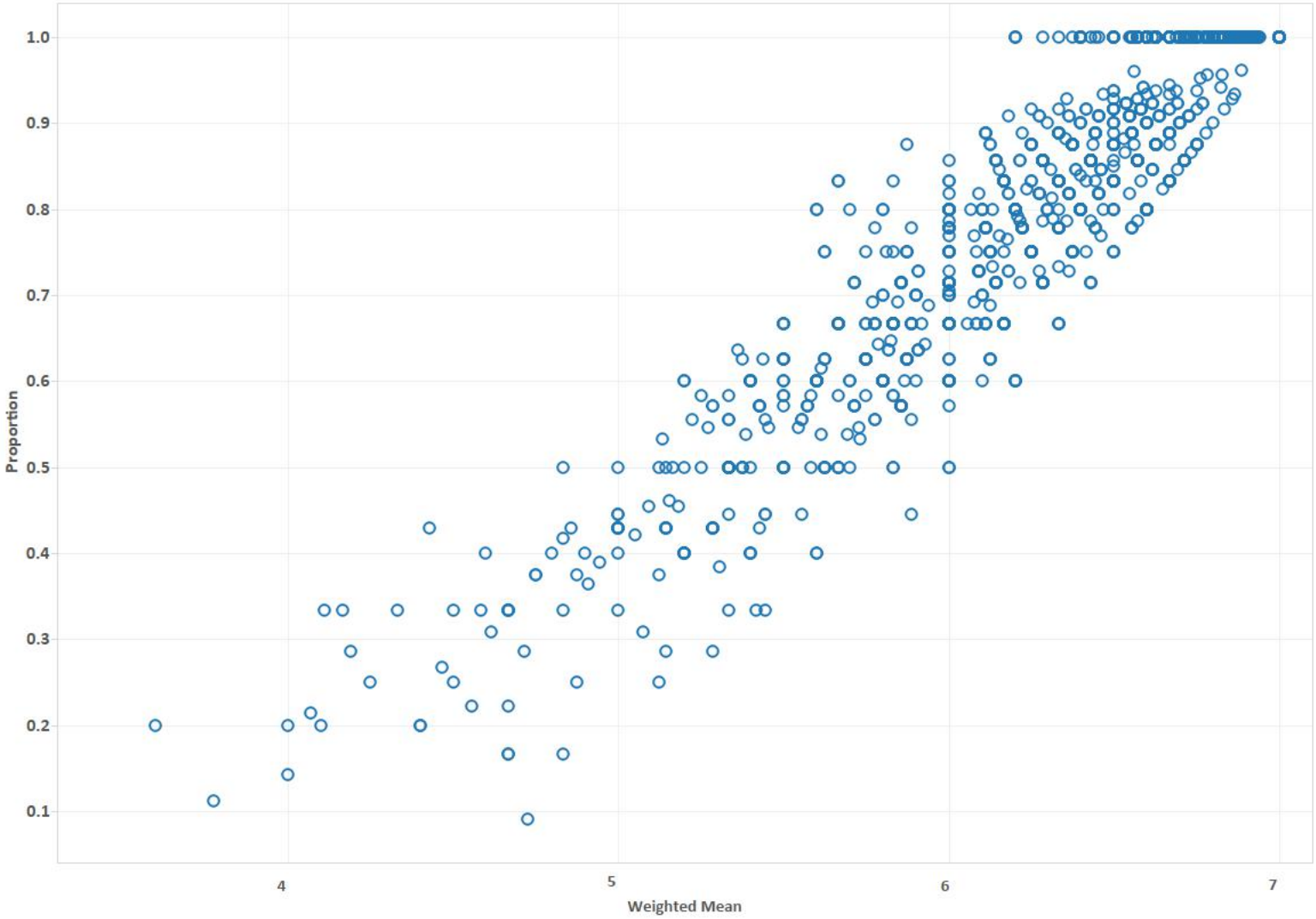


- Means and proportions don't provide the same information
- They will, however, agree at the extremes:
  - When all learners rate an instructor poorly, both mean and proportion will be low
  - When all learners rate an instructor highly, both mean and proportion will be high
  - In between these extremes (where most responses lie) the two will differ

MEAN VS. PROPORTION OF EXCEPTIONAL RESPONSES



MEAN VS. PROPORTION OF (EXCELLENT, EXCEPTIONAL) RESPONSES



- Scatterplots of mean vs. p67 and p7 (each circle represents one course)
- Relationship between mean and p7 is decidedly non-linear, indicating little agreement between these two statistics.
- Relationship between mean and p67 much more linear, indicating greater agreement.

- Are more informative about changes in performance
- Consider the following two **hypothetical** scenarios for two quarters.
- Scenario 1

PSYC 3540	Mean	P45	P12	P1
Q1	3.80	0.6	0	0
Q2	3.60	0.4	0	0

- Scenario 2

PSYC 3540	Mean	P45	P12	P1
Q1	3.80	0.6	0	0
Q2	3.64	0.66	0.34	0.34

1. Survey questions need to be well developed
   - Similar and unambiguous interpretation by all learners
2. Seek input from learners where they actually have the ability and knowledge to provide judgement.
   - Learners may not have the ability to judge whether “The instructor challenged learners to think critically.”
   - They are, however, in a better position to judge whether “The instructor provided clear course expectations (facilitation, assignments, availability)”.

3. Means should never be used in isolation.
4. Ideally, goals should be set in terms of proportions, not means.
   - Goals should be program or course specific
   - Appropriate goals need to incorporate intervals to account for uncertainty