Feinstein School of Education and Human Development
Faculty Meeting
February 9, 2007

1. <u>Welcome (Julie, who introduces Roger)</u>
2. <u>Introduction (Roger introduces meeting)</u>
   a. Focus on assessment and consistency
3. <u>Scoring Issues Overview (Susan)</u>
   a. Some possible key points:
        i. Assessment systems based on human judgment need to develop reliability and validity
       ii. Reliability= the degree to which assessment results are consistent, dependable or repeatable
            1. for example, different scorers viewing the same evidence making the same judgment, or the same scorer making the same judgment about the same evidence on different occasions (maybe show/discuss cartoon on last page of this agenda as an example of a lack of reliability in assessment)
      iii. Reliability is our focus today
       iv. There is no validity without reliability; without reliability, data are not meaningful; judgments aren't fair
4. <u>Scoring Exercise</u>
   a. Participants move to color coded tables that match the color of their folder
   b. Participants score sample of student work on their own, using the rubric provided
5. <u>Small Group Discussion (facilitated by Roger, Susan, Dick, Sue, Julie at tables)</u>
   a. Participants share their scores and justification for their scores at their tables
   b. Discussion questions:
        i. What is the evidence that this piece of student work does or does not meet the standard?
       ii. Does this work sample provide sufficient evidence about student achievement on this standard? Why or why not?
      iii. What further evidence might be helpful to demonstrate achievement of the standard?
6. <u>Large Group Discussion</u>
   a. How reliable are we in our scoring? Calculate inter-rater reliability using show of hands and abbreviated version of inter-rater reliability scoring worksheet (attached)
   b. Discuss:
        i. Based on this activity, is this assessment meaningful?
       ii. What are the implications of this activity for assessment in your department?
      iii. What are the implications for program improvement?
7. <u>Conclusion</u>

R AT E R AGREEMENT SUMMARY SHEET

Student Work Sample	Rubric Score			
	1	2	3	4
A				
B				
C				
D				
E				
F				

Calculating Exact Score Agreement:

1. First, record the number of raters at each score point. Then, circle the highest number for each work sample. Add the numbers in the circles. This is to calculate the number of times the raters were in highest agreement with each other for work samples A-D.   _____

2. Next, multiply the number of raters by the total number of work samples that were scored.   _____

3. Then, divide the number of times the raters were in agreement with each other (Step 1) by the total number of work samples scored (Step 2). This is the percent of exact score agreement for work samples A-D.   _____
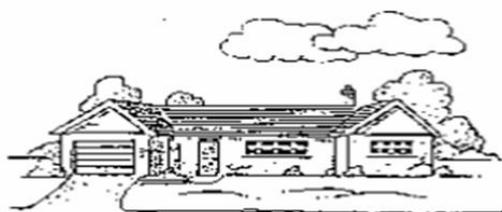
4. Re-calculate the percent of exact score agreement once all six work samples have been scored. (Repeat Steps 1-3.)   _____

YOUR HOME AS SEEN BY

YOU
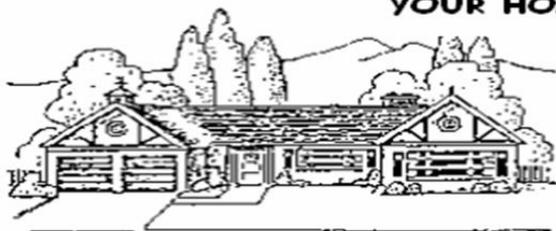
YOUR BUYER

YOUR LENDER

YOUR APPRAISER.

YOUR TAX ASSESSOR

Aspects of the FSEHD Assessment System:

Human Judgment and Reliability

Reliability

- Degree to which assessment results are consistent, dependable, or repeatable
  - Different scorers viewing the same evidence making the same judgment
  - Same scorer making the same judgment about the same evidence on different occasions
- An issue of fairness (validity)

YOUR HOME AS SEEN BY

YOU

YOUR BUYER

YOUR LENDER

YOUR APPRAISER.

YOUR TAX ASSESSOR

How much is enough?

- Required level of reliability for assessment increases as the stakes attached to the assessments increase
  - When assessment-based decisions are important, permanent, or have lasting consequences

Reliability is not inherent

- Reliability must be monitored continuously in all aspects of the design and use of an assessment system
- With low reliability, validity of the system is questionable

How to improve reliability

- Inter-rater reliability dialogue
  - Discussion of students' performances and the ratings that should be applied to those performances
    - Helps establish and clarify standards and ensure that faculty have similar expectations for students.

Discussion questions

- What is the evidence that this piece of student work does or does not meet the standard(s)?
- Does this work sample provide sufficient evidence about student achievement on each standard? Why or why not?
- What further evidence might be helpful to demonstrate achievement of the standard(s)?

Large group discussion

- Based on this activity, is this assessment meaningful?
- What are the implications of this activity for assessment in your department or program?
- What are the implications for program improvement?