THE INSTITUTE FOR Learning

Developing an Instructional Quality Assessment

April 24, 2003
AERA, Chicago

Amy Crosson (LRDC, Pgh),
Brian Junker (CMU, Pgh),
Lindsay C. Mastumura (UCLA),
Lauren Resnick (LRDC, Pgh)

© 2002 University of Pittsburgh

Instructional Quality Assessment (IQA): ***Why?***

- Accountability based on student achievement standards OK as far as it goes…
- Fails to address issue of equitable access to high *instructional quality; intimately related to professional culture in schools*
- *Research tool*: Is long-term improvement in achievement scores related to changes in professional culture?
- *External evaluation tool*: Does PD change the culture?
- *Internal self-study tool*: Clearly define what PD should aim for!

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

What is the IQA?

As an *external* assessment the IQA should provide:

- A performance assessment that measures the quality of teaching at a given point in time

- A school-level index that yields patterns of performance across a sample of classrooms

- A statistically-rigorous tool for evaluating school and district improvement efforts

- An affordable and efficient measure

© 2002 University of Pittsburgh

What is the IQA?

As as an *internal* assessment the IQA should provide:

• Rubrics and criteria charts to guide professional learning about powerful instructional practices

• A performance assessment that tracks improvement in the quality of teaching over time

- teacher self-assessment
- coach diagnosis and planning
- principal assessment of professional development efforts

© 2002 University of Pittsburgh

Components of the IQA

- Observation and interview protocols
- Scoring rubrics
- Rater training program
- Statistical tools to analyze data

This talk is mostly about the external IQA

© 2002 University of Pittsburgh

Development of IQA to date

- Conceptualization; Literature Review
- Scoring rubrics, observation and interview protocols
- Micro-feasibility studies (single classrooms and archival tape)
- Spring 2002 Pre-Pilot Study (12 classrooms, 1 district)
- Revision of rubrics and protocols
- Winter 2003 Pre-Pilot Study (8 classrooms, 2 districts)
- Design and Preparation for Spring 2003 Pilot (48 classrooms, 2districts)
- So far, *math* and *reading comprehension* in *elementary* and *middle grades*.

© 2002 University of Pittsburgh

Issues

- Defining instructional quality
- Data sources
- High versus low-inference observation and rating
- Aggregation rules
- Sampling

© 2002 University of Pittsburgh

Defining Instructional Quality

Principles of Learning

- Academic Rigor
- Accountable Talk[SM]
- Clear Expectations
- Self-management of Learning

© 2002 University of Pittsburgh

Data Sources

- Classroom Visit, including student interviews
- Teacher Interview
- Assignment Portfolios (4 assignments/teacher)
- Principal Interview (*)

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Classroom Visit

- Observation of one full lesson in an identified content area
  - During teacher-led discussion: focus on Accountable Talk℠
  - During independent/group work: focus on Academic Rigor
- Student interviews
  - Focus on Clear Expectations and Self-management of Learning
  - Each rater interviews 2-3 students for approximately 5 minutes
- Contextual information provided by teacher through pre-visitation questionnaire
  - Goal of lesson
  - How lesson fits into instruction over time

© 2002 University of Pittsburgh

Teacher Interview

- Follows lesson visit
- Gain additional contextual information and reflect on lesson with teacher
  - Representativeness of the lesson
  - Modifications made during the lesson
  - Clarification of goals and how this lesson fits into the instructional unit
- Approximately 30 minutes

© 2002 University of Pittsburgh

Collection of Assignments and Student Work

- 4 assignments selected by teacher
  - 2 “recent” assignments
  - 2 “especially challenging”
- For each assignment, 6 samples of student work
  - Sample represents varying performance levels
  - Student work as a lens to understand the assignment-- not rated separately
- Contextual information provided by teacher through cover sheet for each assignment
  - Goal of assignment
  - Description of scoring guides and opportunities for self-assessment

© 2002 University of Pittsburgh

Principal Interview

- Intentions:
  - Gather information about standards, curricula and assessments
  - Focus on professional development opportunities provided for school faculty
  - Approximately 30 minutes
- Not implemented!
- Turns out most the IQA-relevant forms of this info come from teachers...

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

High vs. Low Inference Observation and Rating

- The definition of each Principle of Learning is a rich and detailed feature list
- The "evidence" is buried in
  - real-time social interaction in the classroom
  - off-line student-teacher interaction concerning written assignments
- When these two meet, how do we obtain a reliable, confident 4-category rating?

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Example: Accountable Talk℠

- Talk is essential to learning
- Talk must be accountable

  … to the learning community

  … to knowledge

  … to rigorous thinking

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Accountability to the Learning Community

- Actively participate in classroom talk
- Listen attentively
- Elaborate and build on each other's ideas
- Work to clarify or expand a proposition

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Accountability to Knowledge

- Specific and accurate knowledge
- Appropriate evidence for claims and arguments
- Commitment to getting it right

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Accountability to Rigorous Thinking

- Synthesize several sources of information
- Construct explanations and
  test understanding of concepts
- Formulate conjectures and hypotheses
- Employ generally accepted standards of reasoning
- Challenge the quality of evidence and reasoning

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

To help raters distill this information

- Limit the number of ratings that are needed
- Focus raters' attention on the essential features, using "checklists"
- Using checklists *only as guides*, raters aggregate observations into a more reliable/confident rating scale score.

© 2002 University of Pittsburgh

Limit the number of ratings needed:

Accountable Talk Dimensions

- How effectively did the lesson talk build accountability to the learning community?
  - Participation
  - Linking student contributions
- How effectively did the lesson talk build accountability to knowledge?
- How effectively did the lesson talk build accountability to rigorous thinking?

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

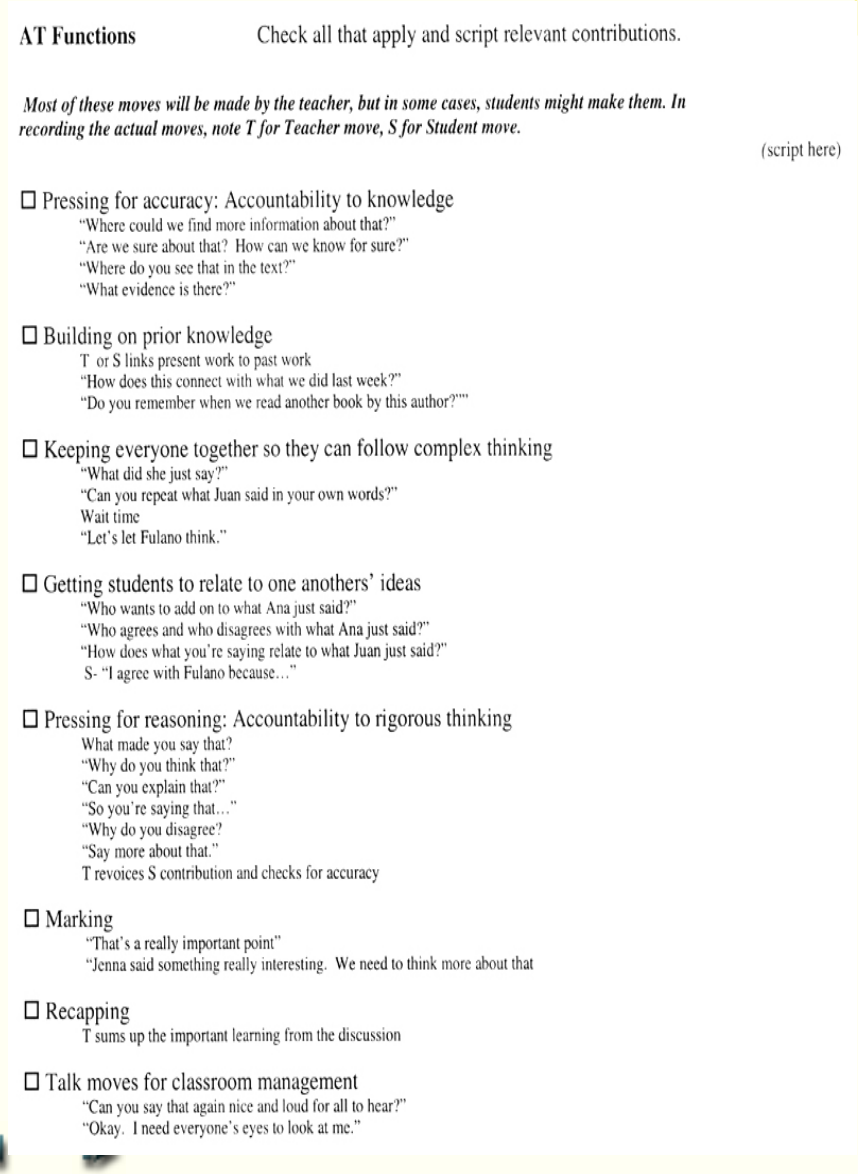
Focus raters' attention using "Checklists"

- Here the theory of Accountable Talk is helpful; there are many "talk moves" that are partial indicators of AT.
  - Pressing for accuracy…
  - Building on prior knowledge…
  - Keeping everyone together…
  - etc.

© 2002 University of Pittsburgh

AT Functions structure observations

- Raters script talk exchanges that serve various talk functions
- Both teacher and student contributions



AT Functions Check all that apply and script relevant contributions.

Most of these moves will be made by the teacher, but in some cases, students might make them. In recording the actual moves, note T for Teacher move, S for Student move.

(script here)

☐ Pressing for accuracy: Accountability to knowledge
"Where could we find more information about that?"
"Are we sure about that? How can we know for sure?"
"Where do you see that in the text?"
"What evidence is there?"

☐ Building on prior knowledge
T or S links present work to past work
"How does this connect with what we did last week?"
"Do you remember when we read another book by this author?""

☐ Keeping everyone together so they can follow complex thinking
"What did she just say?"
"Can you repeat what Juan said in your own words?"
Wait time
"Let's let Fulano think."

☐ Getting students to relate to one anothers' ideas
"Who wants to add on to what Ana just said?"
"Who agrees and who disagrees with what Ana just said?"
"How does what you're saying relate to what Juan just said?"
S- "I agree with Fulano because…"

☐ Pressing for reasoning: Accountability to rigorous thinking
What made you say that?
"Why do you think that?"
"Can you explain that?"
"So you're saying that…"
"Why do you disagree?
"Say more about that."
T revoices S contribution and checks for accuracy

☐ Marking
"That's a really important point"
"Jenna said something really interesting. We need to think more about that

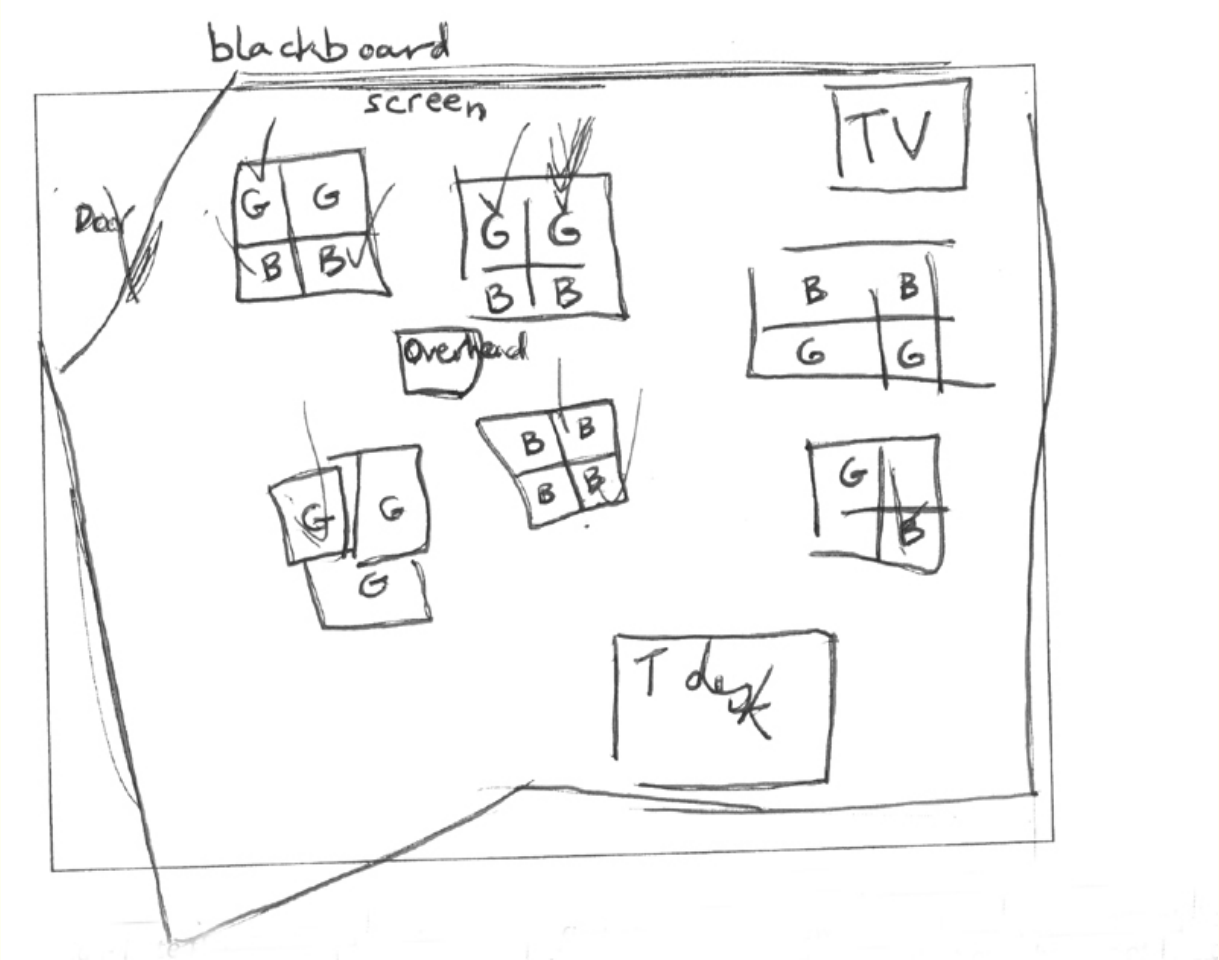
☐ Recapping
T sums up the important learning from the discussion

☐ Talk moves for classroom management
"Can you say that again nice and loud for all to hear?"
"Okay. I need everyone's eyes to look at me."

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Classroom map



© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Raters use checklists etc. to guide rubric scoring

- 1-2 Rubrics for each dimension
- AT functions checklist, scripting, etc. are used to guide rubric scoring

© 2002 University of Pittsburgh

Example

- **Dimension:** *How effectively did the lesson talk build accountability to the learning community?*

- **Rubric:** Linking student contributions: Did speakers' contributions link to and build on each other? (i.e., Was there "local coherence" during the discussion?)

- **Talk functions to look at:**
  - ❑ Keeping everyone together so they can follow complex thinking
  - ❑ Getting students to relate to one anothers' ideas

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

A. How effectively did the lesson-talk build <u>Accountability to the Learning Community</u>?

A1. Participation: Was there widespread participation in teacher-facilitated discussion?

4	Over 50% of the students participated consistently throughout the discussion.
3	25-50% of the students participated consistently in the discussion OR over 50% of the students participated minimally.
2	25-50% of the students participated minimally in the discussion (i.e, they contributed only once).
1	Less than 25% of the students participated in the discussion.

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

- **A2. Linking student contributions:** Did speakers' contributions link to and build on each other? (i.e., Was there "local coherence" during the discussion?)

4	At numerous points during the discussion, the teacher/student explicitly connects others' contributions and shows how ideas/positions shared during the discussion relate to each other.
3	At a few points during the discussion, the teacher /student links others' contributions to each other and shows how ideas/positions relate to each other.
2	At a few points during the discussion, the teacher/student links others' contributions to each other.
1	Teacher/student does not make any effort to link others' contributions.

© 2002 University of Pittsburgh

Academic Rigor Dimensions

- Engagement with core academic content
- Application of complex thinking skills
- Elaborateness of student response
- Focus of teacher's goals on complex thinking
- Focus of expectations for quality work on complex thinking
- Quality of feedback to students during activities
- Amount of feedback to students during activities

© 2002 University of Pittsburgh

CE/SML Dimensions

- Students' understanding of expectations
- Students' use of criteria to judge work
- Students' use of criteria to revise work
- Access to expectations/grading criteria
- Clarity of assignment directions
- Clarity and detail of grading criteria

© 2002 University of Pittsburgh

Spring 2002 Pre-Pilot

- **Three** Elementary/Middle Schools (K-8)
- **Four** Classrooms per School
  - 5 Mathematics Lessons: Grades 1, 2, 3, 7
  - 7 Reading Comp. Lessons: Grades 1, 2, 3, 6, 8
- Brief Teacher Interviews and Questionnaires
- **Two** Observers per Classroom (in rotating pairs)
  - Lesson Observation
  - Student Interviews
- Portfolios of Assignments from Teachers (Incl. sample student work)
- Rating Forms for
  - Academic Rigor (AR): Math and Reading Comp
  - Accountable Talk (AT)
  - Clear Expectations / Self-Mgmt of Learning

© 2002 University of Pittsburgh

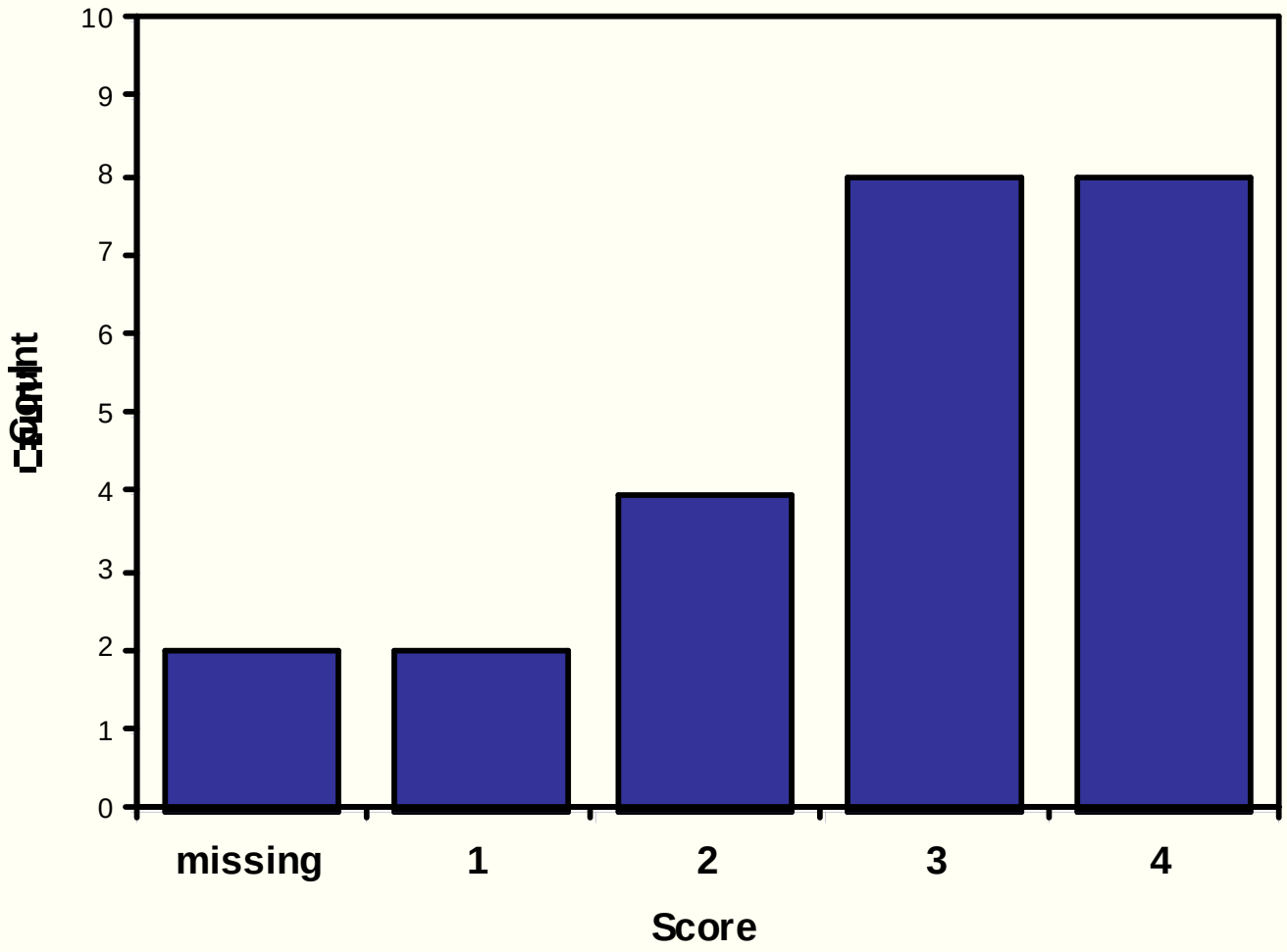
Purpose of Pilot Study

- How do ***observation protocols*** and ***scoring procedures*** *need to be refined*?
-
- Which ***rubrics*** *did not hold up well* for classroom observation and assignment rating?
-
- Which ***rubrics*** *seemed to yield valid information* about the quality of instruction?

- How large were the ***variance components*** and how does this impact *design of full pilot(s)* in 2003?

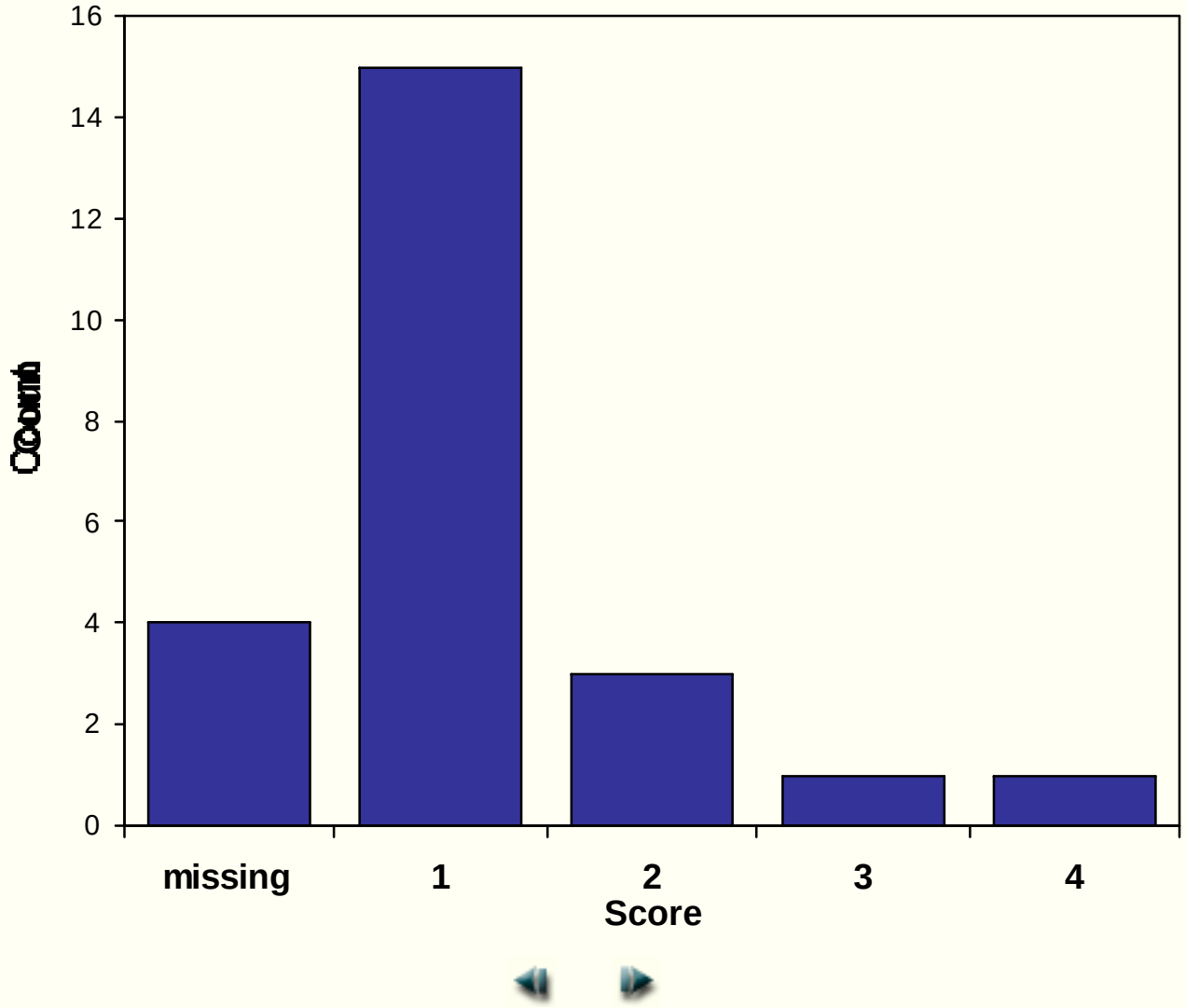
© 2002 University of Pittsburgh

A. Accountability to the Learning Community: A1. Participation



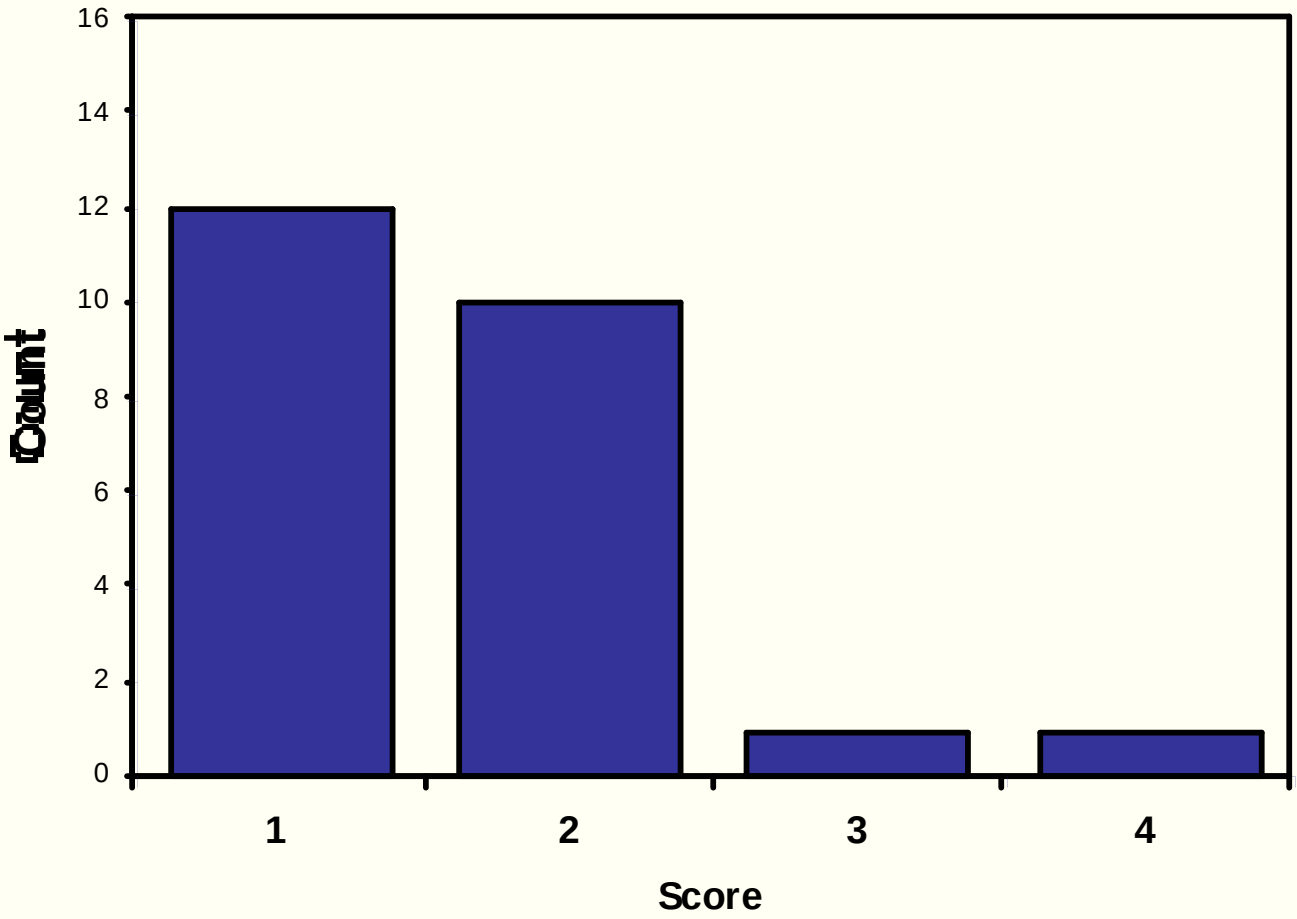
© 2002 University of Pittsburgh

A2. Linking Contributions
Teacher performance



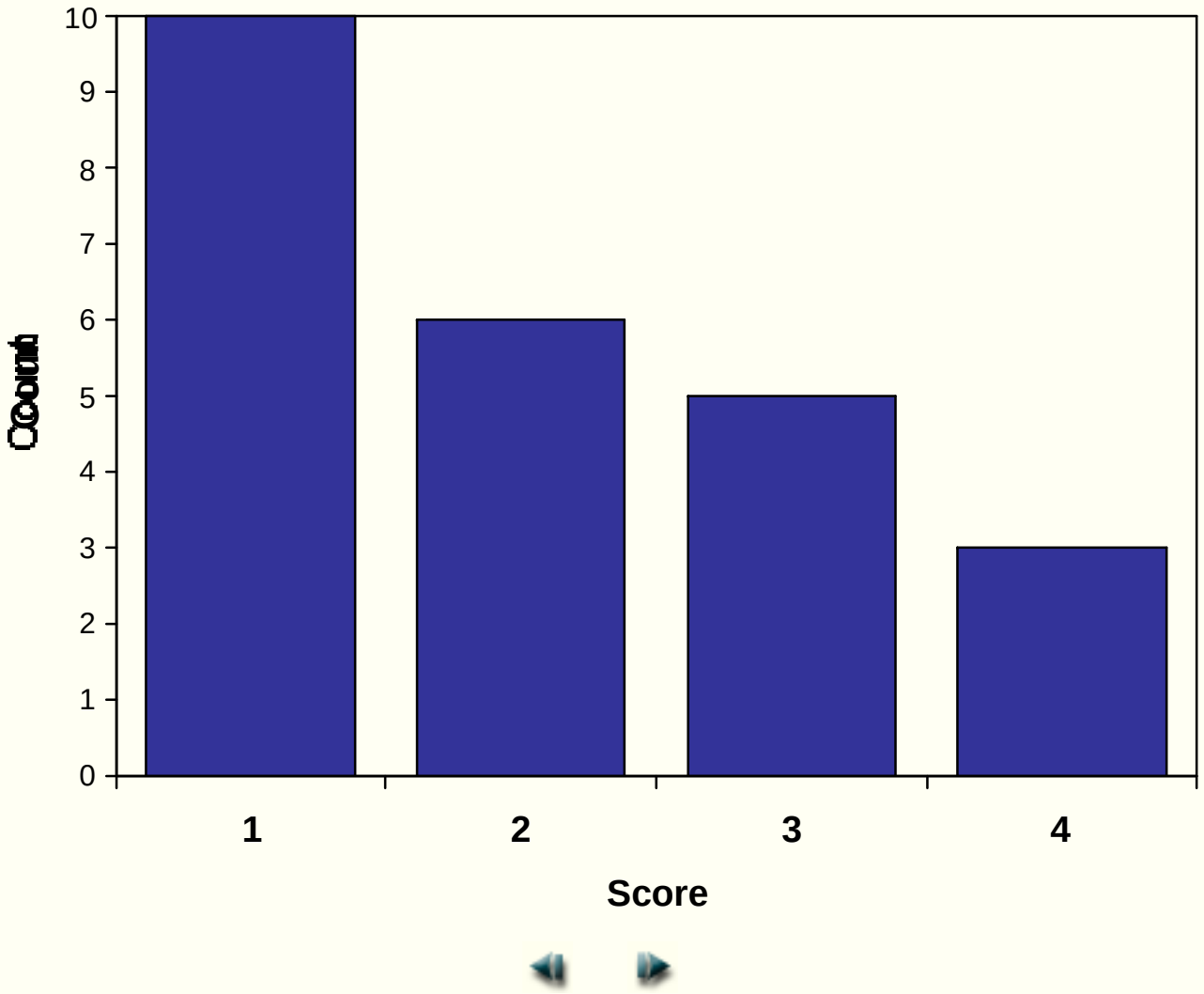
© 2002 University of Pittsburgh

A2. Linking Contributions
Student performance



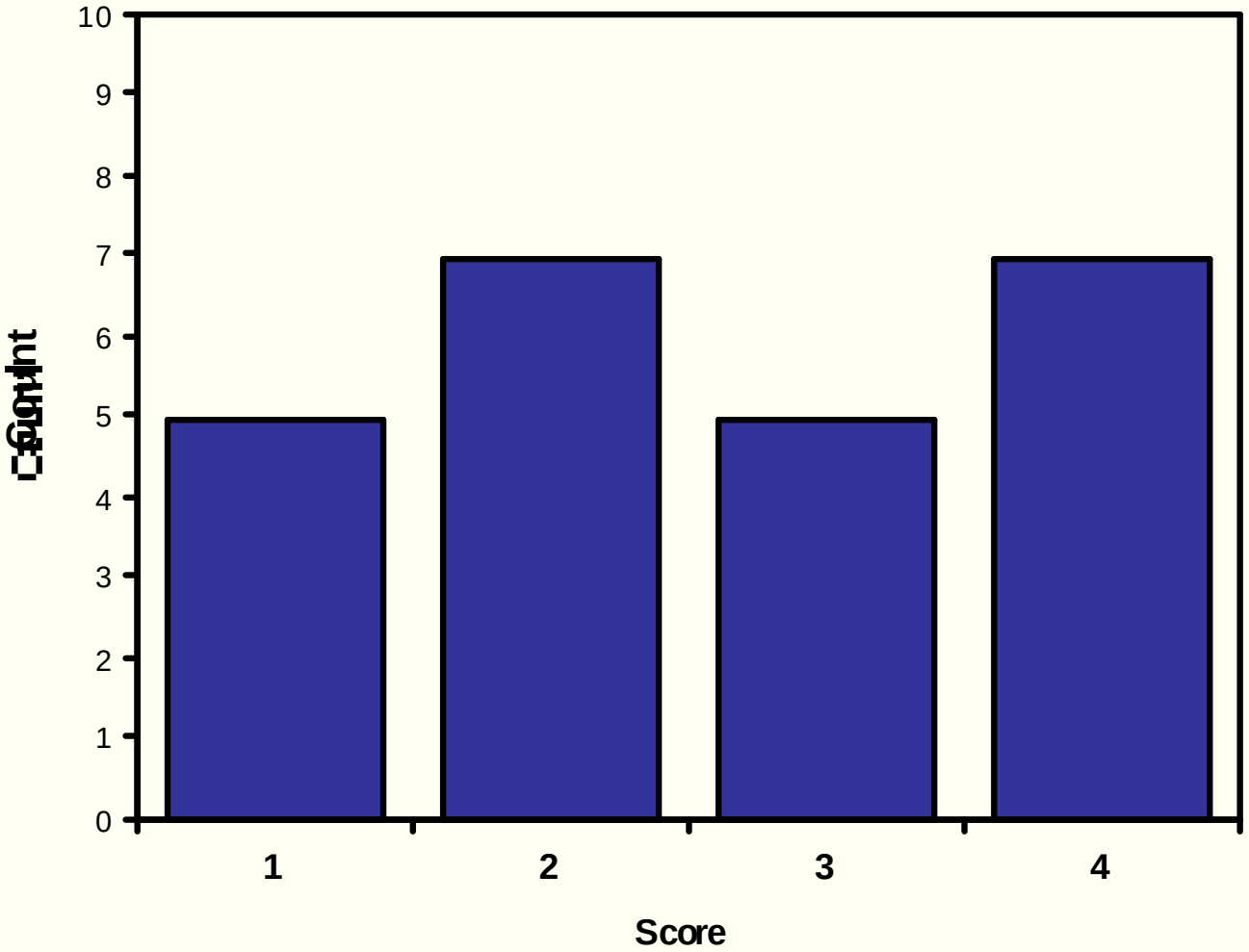
© 2002 University of Pittsburgh

B. Accountability to Knowledge
B1. Pressing with evidence



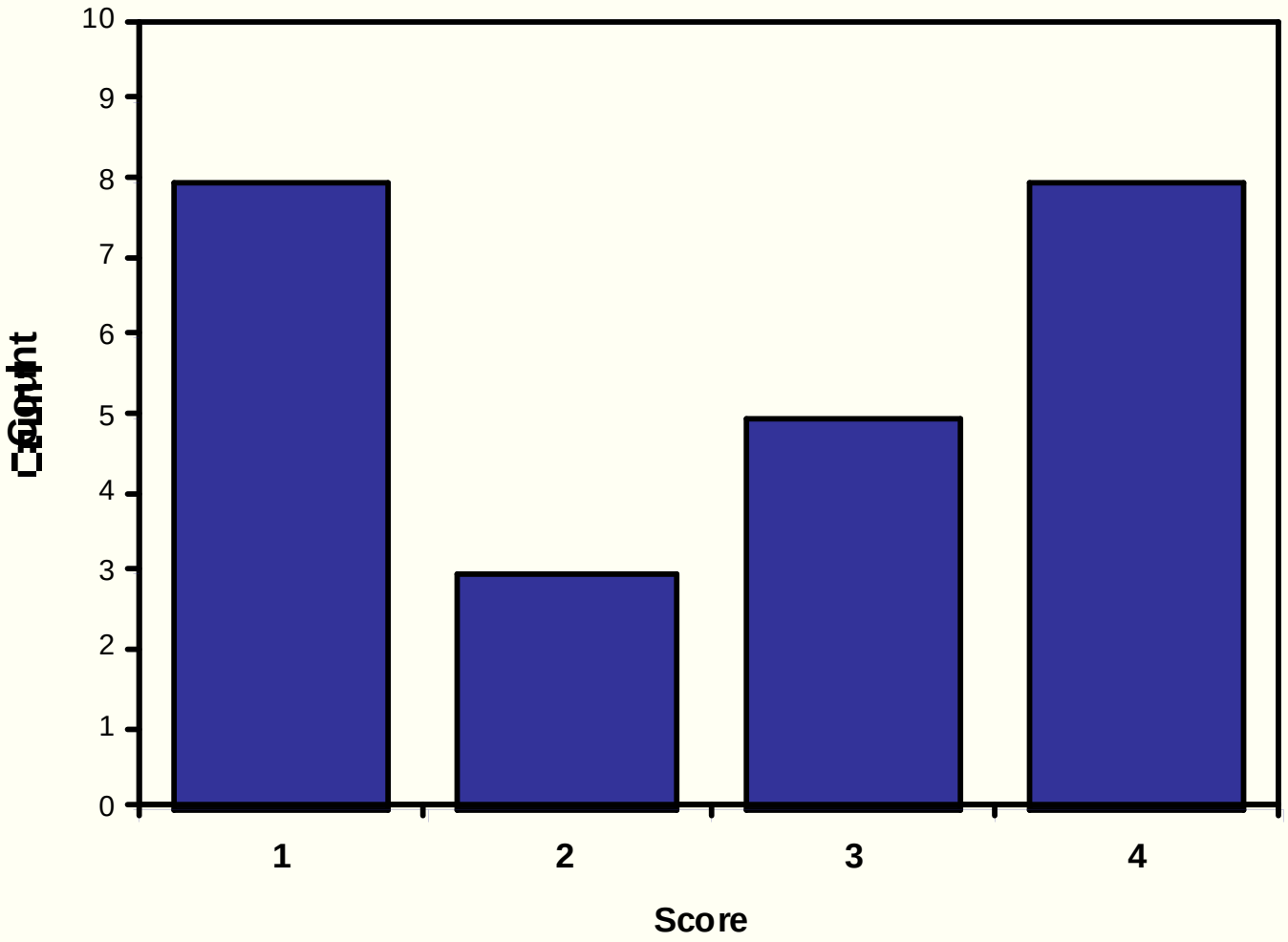
© 2002 University of Pittsburgh

B. Accountability to Knowledge
B2. Supporting with evidence



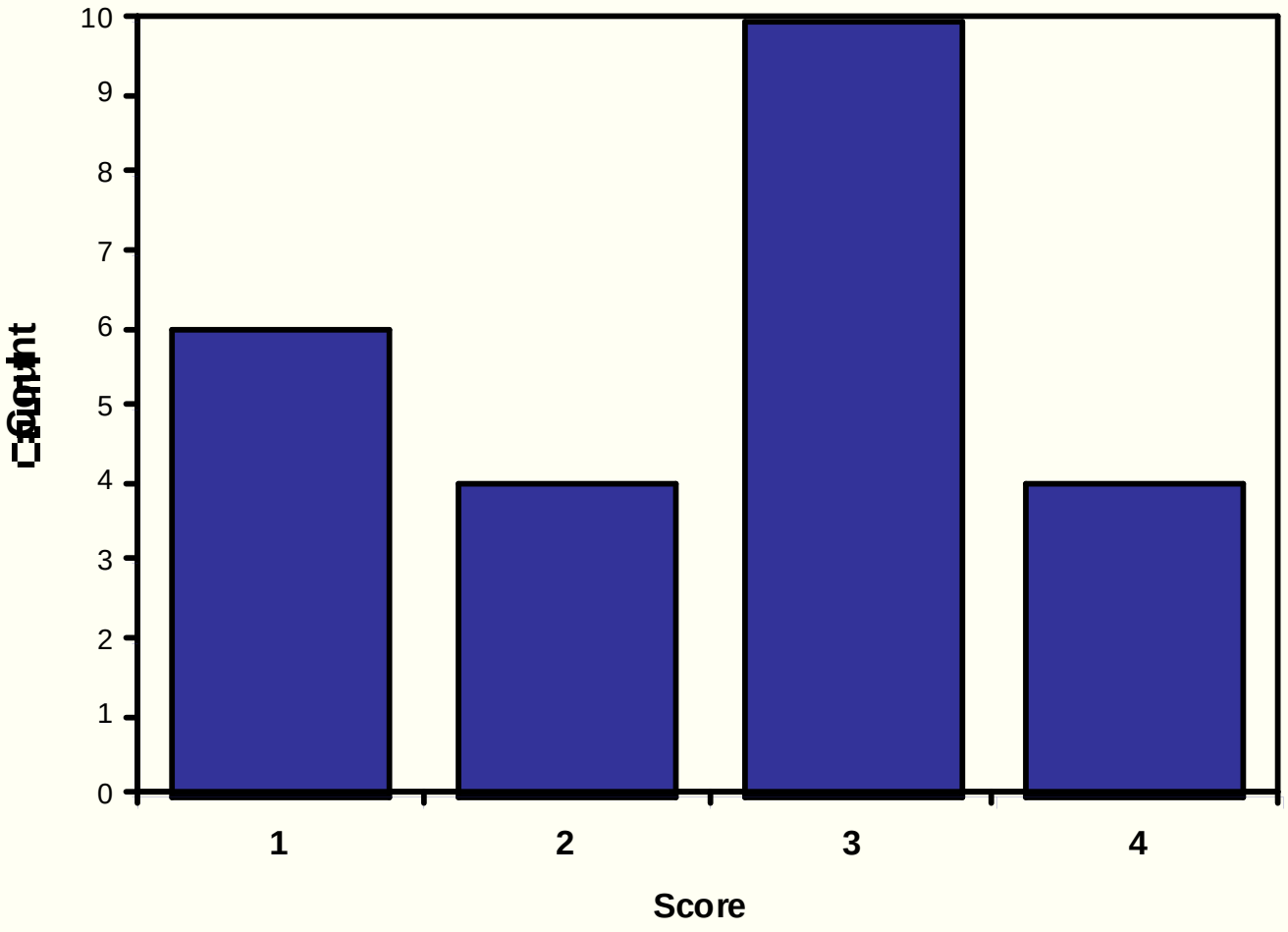
© 2002 University of Pittsburgh

C. Accountability to Rigorous Thinking
C1. Pressing of thinking



© 2002 University of Pittsburgh

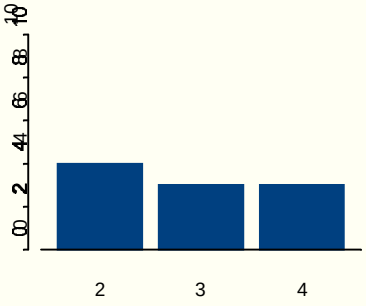
C. Accountability to Rigorous Thinking
C2. Explaining of thinking



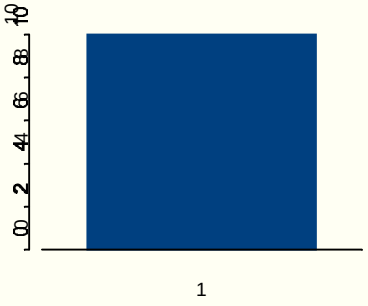
© 2002 University of Pittsburgh

AT Math Initial Ratings

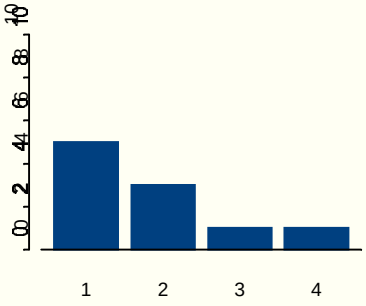
ATD4.I1



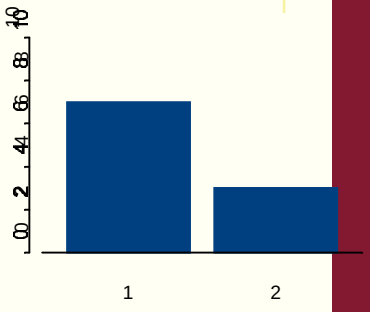
ATD4.I2



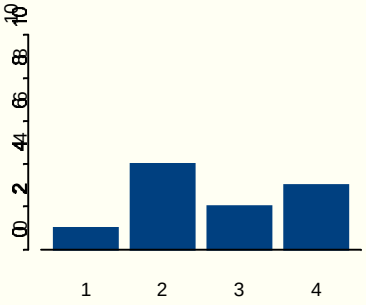
ATD4.I3



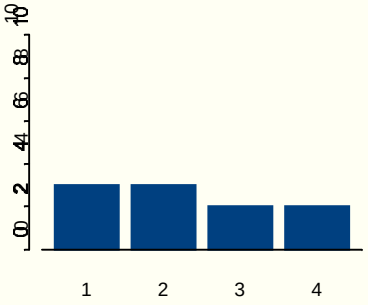
ATD4.I4



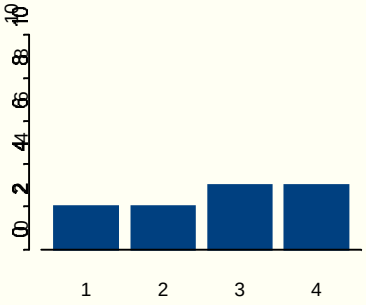
ATD4.I5



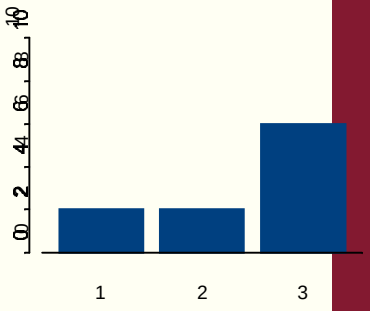
ATD4.I6



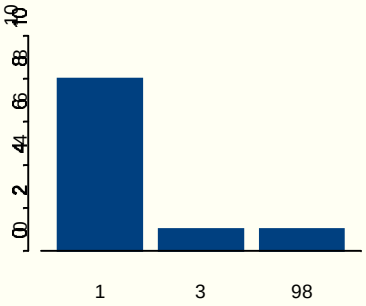
ATD4.I7



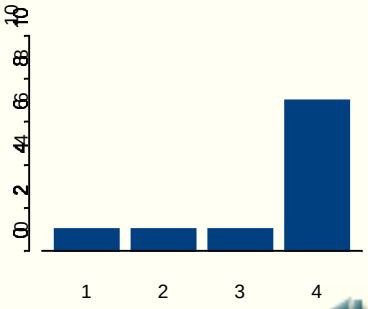
ATD4.I8



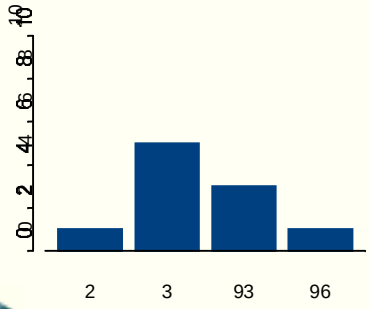
ATD4.I9



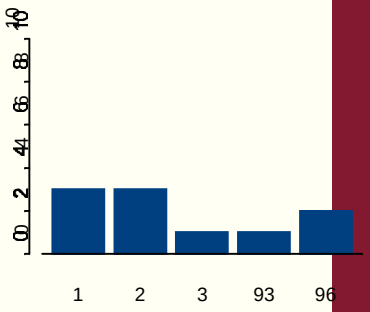
ATD4.I10



ATD4.I11



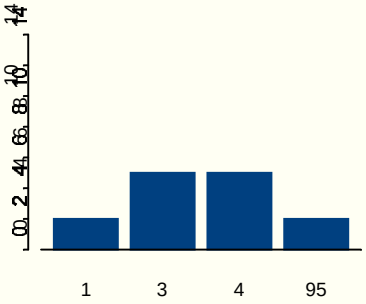
ATD4.I12



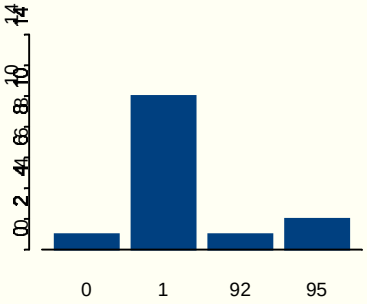
© 2002 University of Pittsburgh

AT Reading Comp Initial Ratings

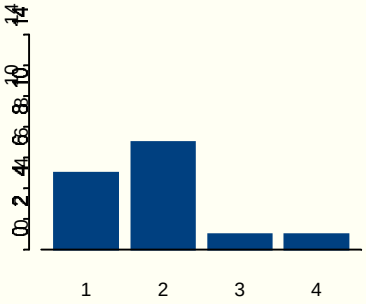
ATD4.I1



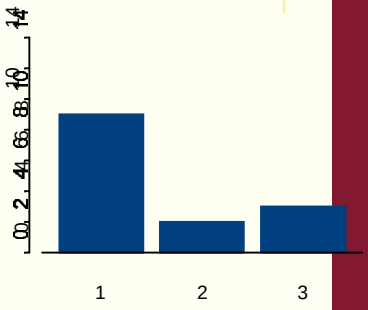
ATD4.I2



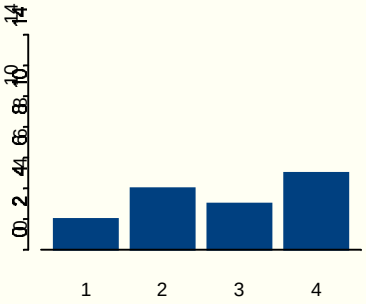
ATD4.I3



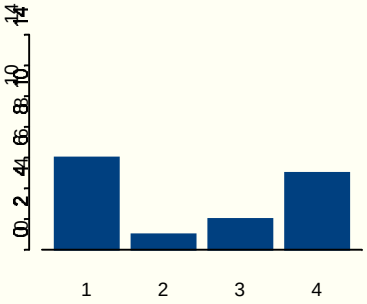
ATD4.I4



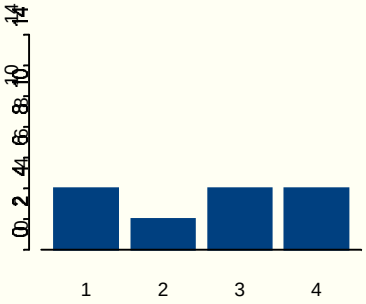
ATD4.I5



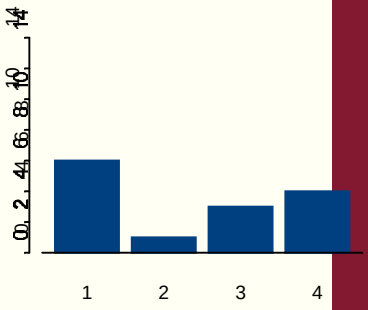
ATD4.I6



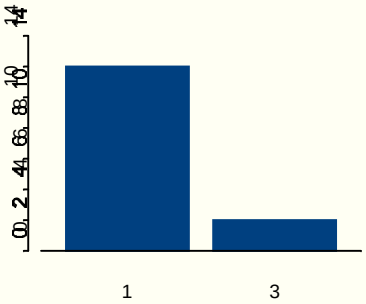
ATD4.I7



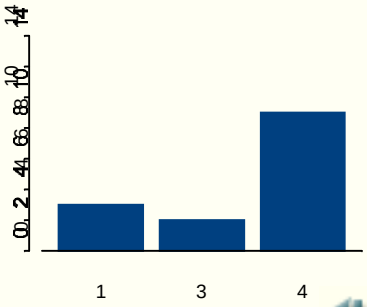
ATD4.I8



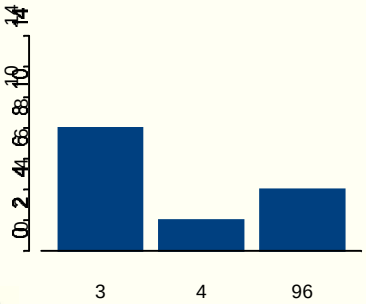
ATD4.I9



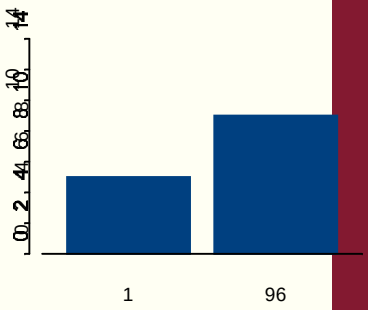
ATD4.I10



ATD4.I11



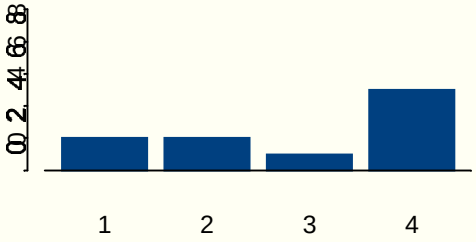
ATD4.I12



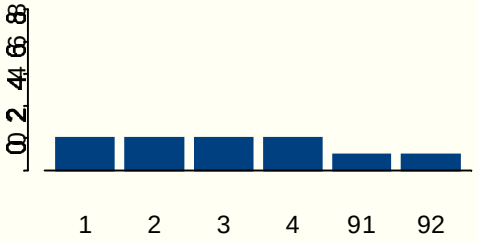
© 2002 University of Pittsburgh

AR Math Initial Ratings

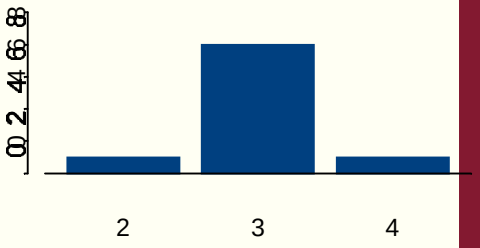
AR.I1



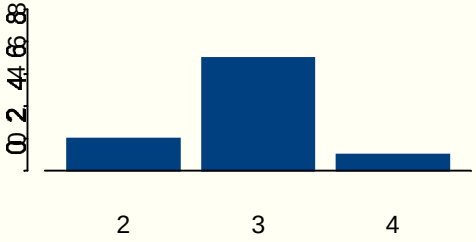
AR.I2C



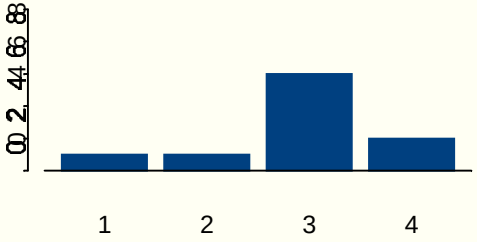
AR.I3



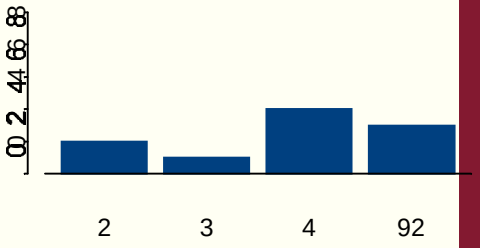
AR.I4



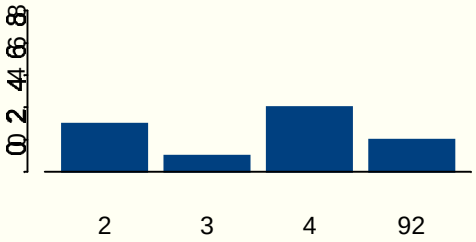
AR.I6



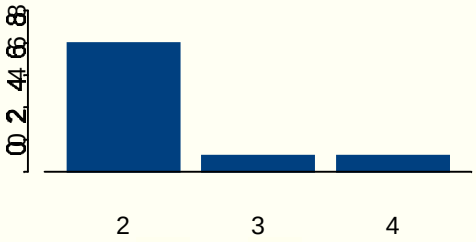
AR.I7



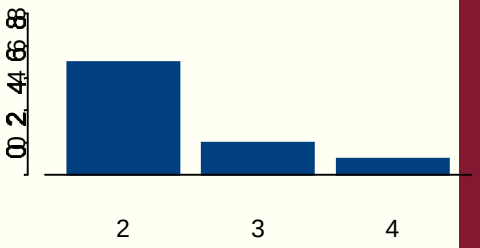
AR.I8



AR.I9



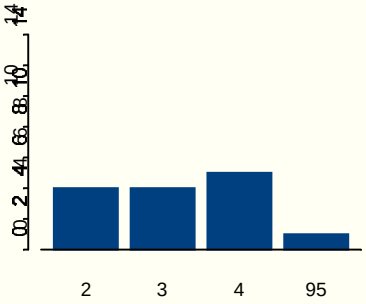
AR.I10



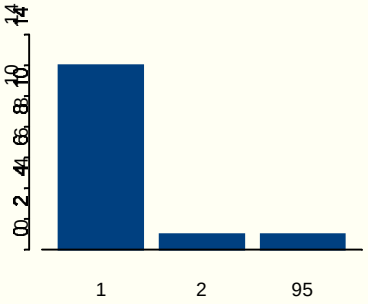
© 2002 University of Pittsburgh

AR Reading Comp Initial Ratings

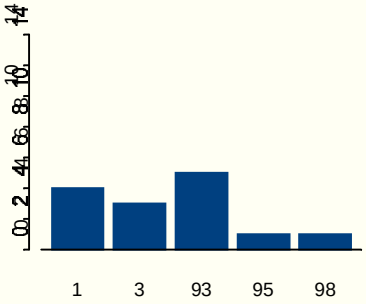
AR.I1



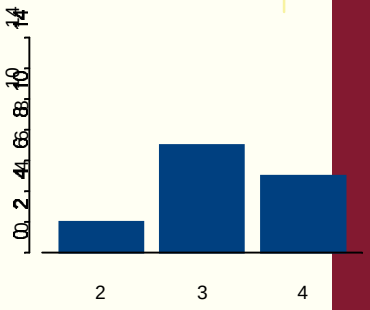
AR.I2C



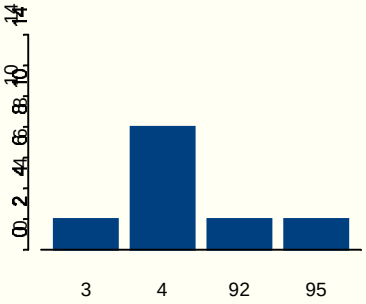
AR.I2D



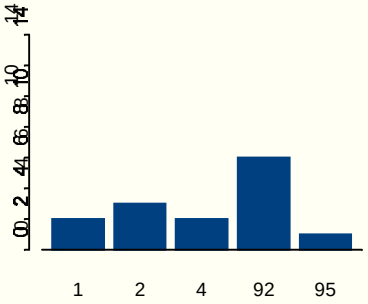
AR.I3



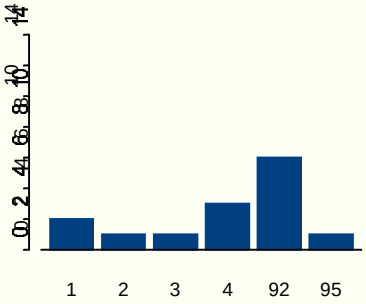
AR.I5



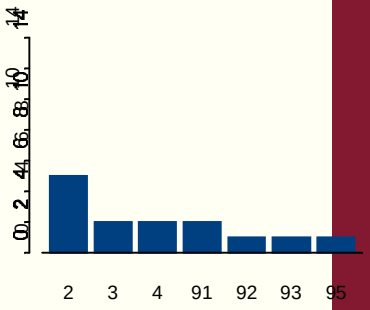
AR.I6



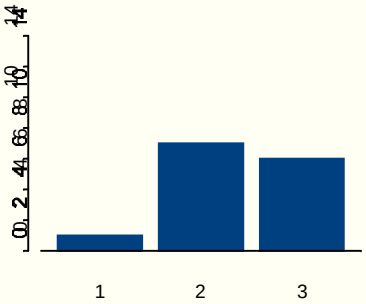
AR.I7



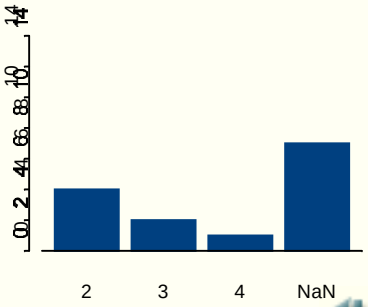
AR.I8



AR.I9



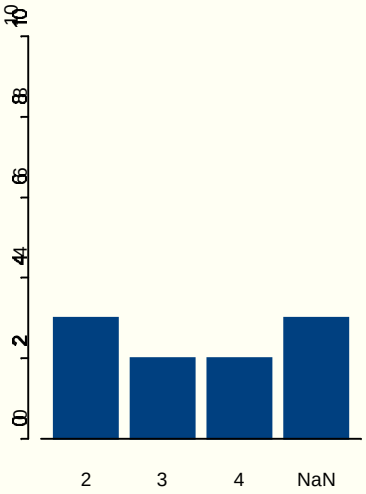
AR.10



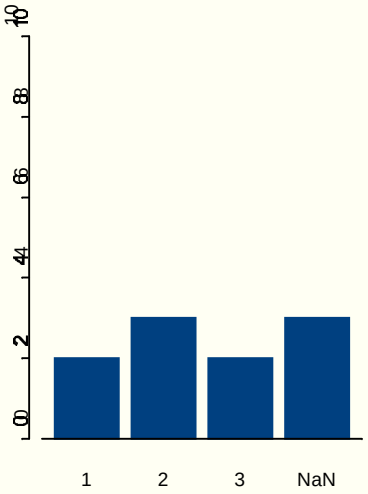
© 2002 University of Pittsburgh

CE Math Initial Ratings

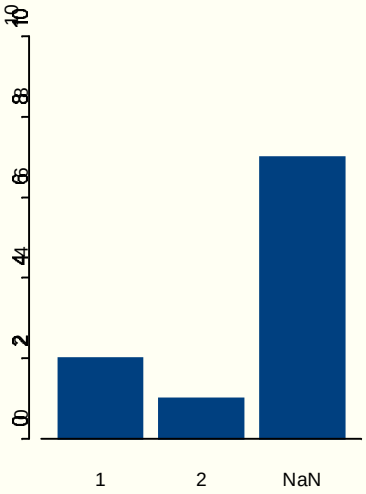
CE.LT12



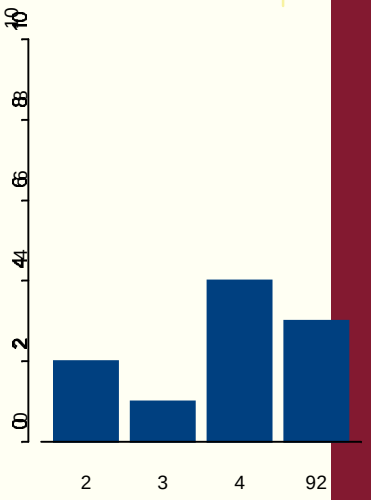
CE.LT13



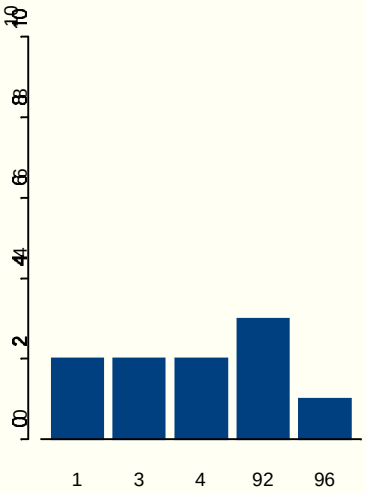
CE.LT14



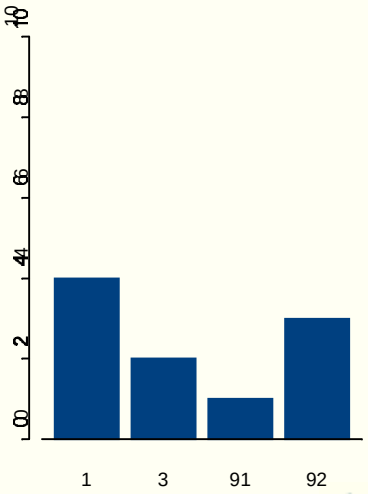
CE.LTI15



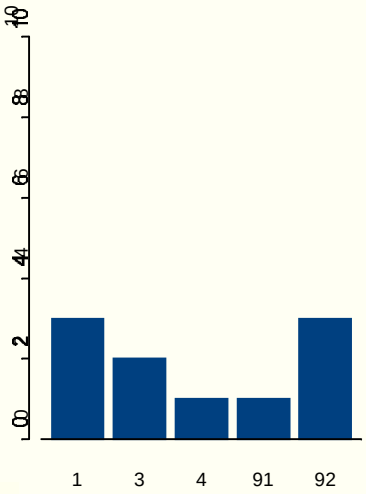
CE.LTI18



CE.LTI19



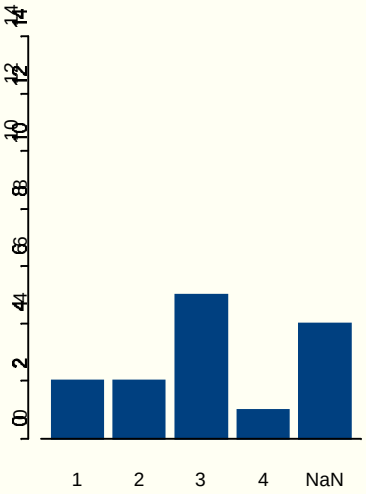
CE.LTI20



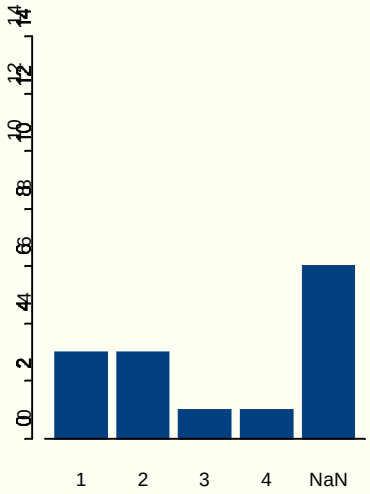
© 2002 University of Pittsburgh

CE Reading Comp Initial Ratings

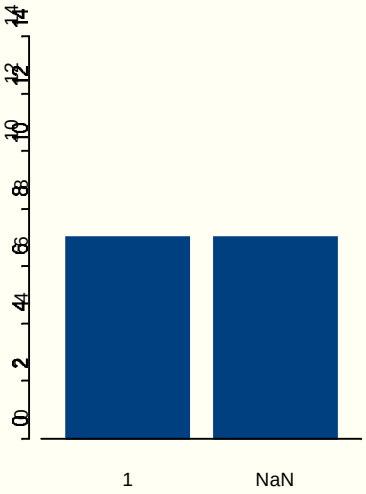
CE.LT12



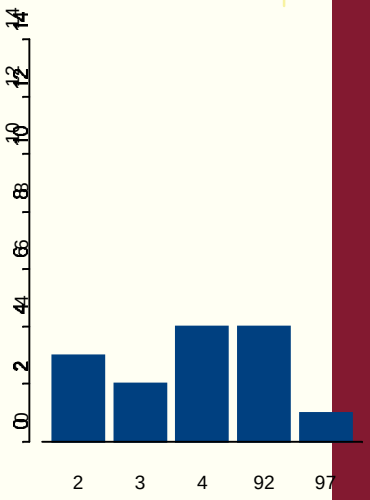
CE.LT13



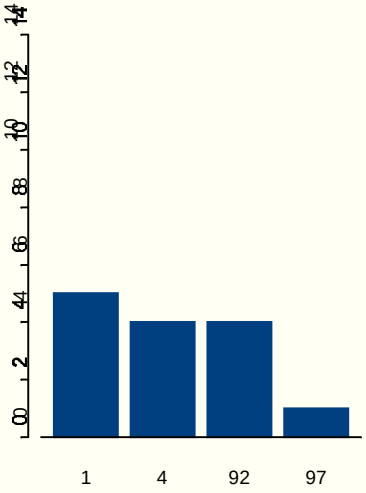
CE.LT14



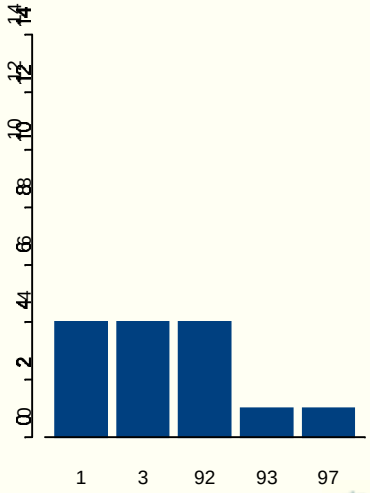
CE.LTI15



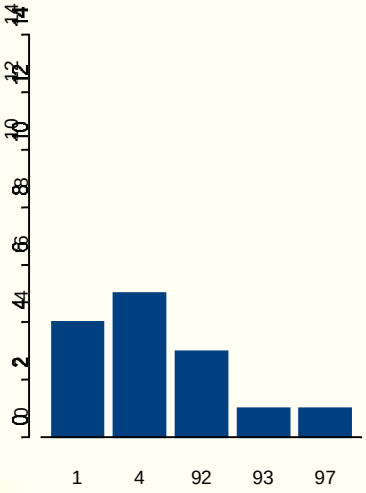
CE.LTI18



CE.LTI19

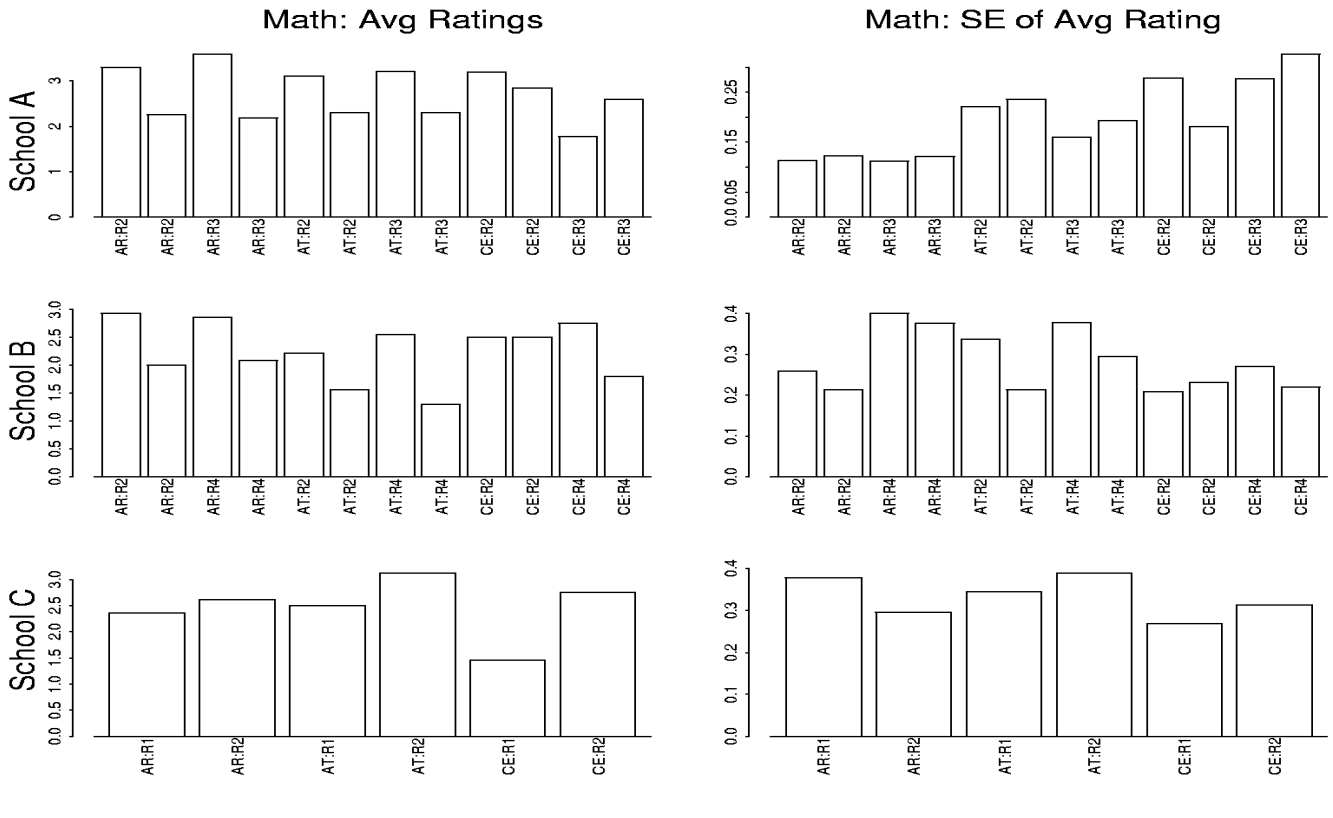


CE.LTI20



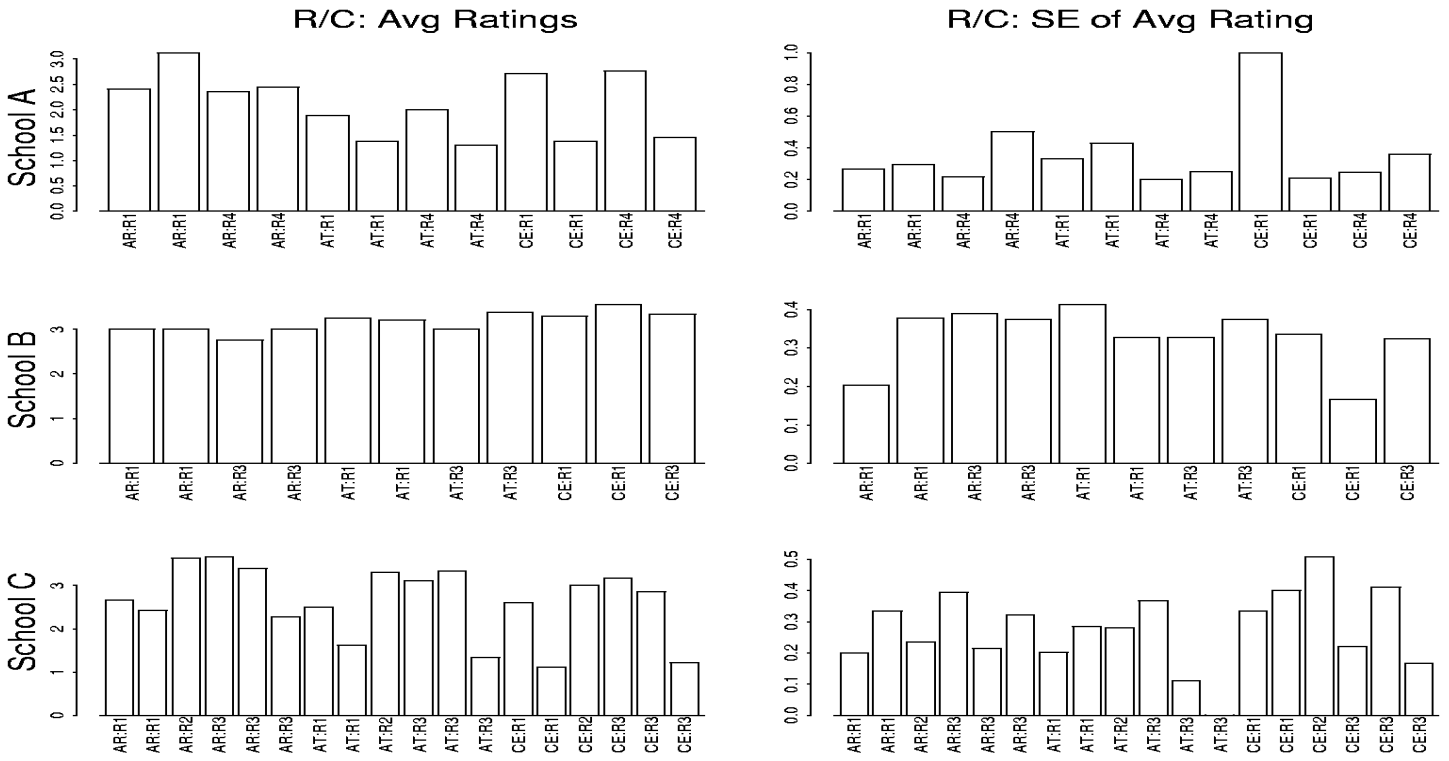
© 2002 University of Pittsburgh

Avg Ratings per rater, class, PoL: Mathematics



© 2002 University of Pittsburgh

Avg ratings per rater, class, PoL: Reading Comprehension



© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Variance Components Analysis: AT

Source of Variation	Estimated Variance Component	Percentage of Total Variance
Teacher:School	.332	41
School	(negl.)	--
Rubric	.164	20
Rater	.036	4
Error	.275	34

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Other Analyses

- Most other variance components analyses could not give useful rater reliabilities, because of
  - small sample size
  - missing data
- Other descriptive summaries were generated
  - crosstabs for rater agreement
  - marginal counts in each category for each question
  - correlations between apparently related rubrics
- Extensive documenations of missing ratings, other problems with the rubrics

© 2002 University of Pittsburgh

Fall 2002 - Winter 2003

- Revision of rubrics
  - Quantitative analysis
  - Qualitative analysis
  - New pre-pilot try-out in Western Pennsylvania (8 classrooms in 2 school districts)
- Development of rater training program
- Planning for Spring 2003 Pilot Study

© 2002 University of Pittsburgh

Some Findings: Protocols and Procedures

- Observers did not have time for both student interviews and classroom observation.
  - Affects AT, CE/SML
  - Supplement interviews with survey for "whole" class
  - Shorten and simplify student interview protocol
- Students in Lower Grades cannot articulate CE without concrete examples.
  - Affects CE/SML
  - Upper grades do not have this difficulty
  - Interviews in lower grades with specific examples of rubrics, student work, etc.

© 2002 University of Pittsburgh

Some Findings: Qualitative Reliability / Validity

- AR/Math: Some rubrics distinguish levels by *type of task* rather than by *challenge of task*.
- AR: **Rigor of Teacher Goals** *difficult to rate* when goals are vaguely stated.
- CE: **Clarity of Task Directions** had *low inter-rater reliability*.
- CE: **Access to Grading Criteria** had *high inter-rater reliability* and either *very high* or *very low ratings*.
- AT: **Teacher Encourages Students to Build on Each Other** and **Press by All Participants** had *high inter-rater reliability*.

© 2002 University of Pittsburgh

Spring 2003 Pilot Study

- May (!!!) 2003
- 24 classrooms (6 schools x 4 classes/school) in each of 2 districts
- Lower/ vs. upper elem., and math vs. R/C
- Schools stratified by achievement scores
- Two newly-trained raters plus IQA staff monitor in each room

© 2002 University of Pittsburgh

Rater Training Program

- Raters recruited from universities near the participating school districts
- Separate sets of raters will be trained for classroom obs, vs rating assignment portfolios
- Test of the "tranferrability" of IQA rating procedures and concepts to "IQA novices".
- Necessary! Our staff cannot handle 48 classrooms alone...

© 2002 University of Pittsburgh

Research Questions For Spring 2003 Pilot

- The districts are similar in all ways except how long they have been studying the Principles of Learning. Can the IQA tell?
- Can grad students and professions in education, not familiar with IQA, be trained to identify and score evidence for the IQA?
- How variable are IQA scores across a district?
- To what extent do IQA scores correlate with state achievement data.

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

The End

© 2002 University of Pittsburgh

THE INSTITUTE FOR Learning

Design Principles: High Performance District

- Commitment to concept of effort-based intelligence
- Focus at every level on classroom instruction, including core principles of learning & teaching
- Nested learning communities
  - Two-way accountability
  - Everyone a learner
- Evidence-based planning and action
- Continuing professional development, based in schools and linked to instructional program
- Coherent standards, curriculum, assessment, & professional development

Student Learning

- Accountability assessments
- Other district performance assessments
- Moderated student grades

Instructional Quality

- Principles of Learning visible in classroom teaching
- Rigorous curriculum and instruction for all students
- High quality of student written & oral work samples

© 2002 University of Pittsburgh