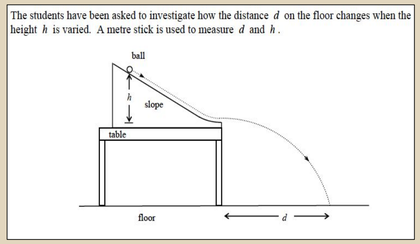
Investigating Student Understanding of Measurement Uncertainty

River Ward*, Trevor Franklin*, Marcos (Danny) Caballero*, and Rachel Henderson*

*MSU Department of Physics and Astronomy

Physics Measurement Questionnaire (PMQ)[1]

Open-ended assessment tool intended on measuring concepts surrounding measurement uncertainty.



Probes of Interest

Repeating Measurements (RD)

Using Repeated Measurements (UR)

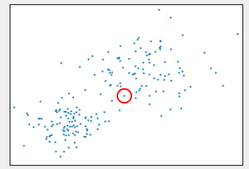
Same Mean with Different Spread (SMDS)

Different Mean with Same Spread (DMSS)

Point-Like Reasoning

Focuses on individual measurements

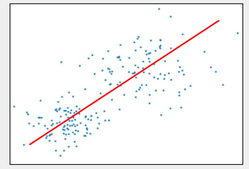
Ignores the presence of statistical error



Set-Like Reasoning

Considers all data together as a whole

Appropriately uses mean and uncertainty

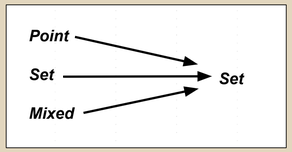


DATA Lab

Design, Analysis, Tools and Apprenticeship Laboratory

- DATA Lab is an algebra-based introductory laboratory course sequence for non-physics majors.
- Transformation was motivated by the AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum.

Ideal shift toward set-like reasoning



- 1188 students responded to the PMQ
- 655 students enrolled in DATA Lab I and 533 enrolled in DATA LAb II

Methods

Coding Procedure

- Two researchers independently code the same subset of student responses.
- An Inter-Rater Reliability (IRR) is calculated and researchers reconcile differences in interpretation.
- Repeat this process until the IRR levels out and overall coding differences are reduced.
- Independently code the remainder of the dataset for that specific probe.

	RD	UR	SMDS	DMSS
Cohen's Kappa	0.60	0.76	0.69	0.56

Final IRR scores for each probe.[2]

Mixed-Like Reasoning

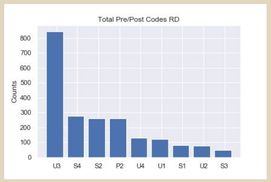
- Codebook developed by researchers at the University of Colorado Boulder.[4]
- Responses can have multiple subcodes associated with them.
- Responses that were not ***purely*** point-like or set-like were reported as "mixed". This includes answers both point-like and set-like or neither.[3]

S1-Simply "average", or names reported value as average	States things like "I averaged," "do the average," "average is best," or "it is the average," but does not elaborate along the lines of the other codes. Includes statements that simply say what the reported value is.
S2-Why average is useful	…reporting the average is best, because (in general) it accounts for fluctuations or errors, or because it predicts future measurements.
S3-Why average is appropriate in this case	…reporting the average is best because all of this data matters, or because the spread of this data is small enough. Includes reporting all data as well as the average. Does not include "it is the correct thing to do" (see S7)

Excerpt from the UR probe codebook

Conclusions

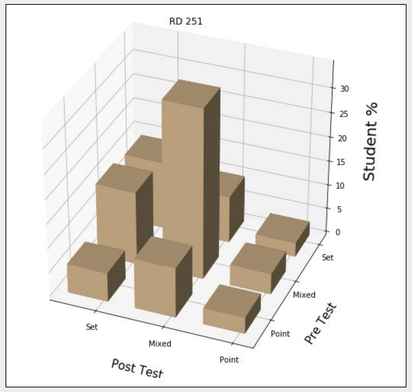
- Paradigm shifts vary significantly depending on probe.
- Assessment tools may have room for improvement.
  - While RD has significant shifts from point-like to set-like, many students remain mixed-like.
  - UR is not useful for measuring student learning.
  - SMDS and DMSS seem to probe into knowledge about mean and spread held before instruction.



Future Work

- Perform a longitudinal analysis of students who took both DATA Lab courses.
- Submit a fully detailed analysis to *The Physics Teacher*.

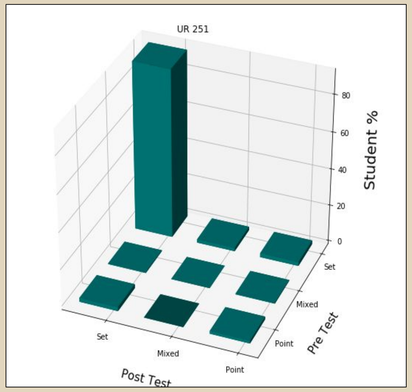
RD - Repeating Measurements



- Measures knowledge that multiple trials are needed to extract meaningful information from an experiment.
- 34% of students remain as mixed-like
- Similar levels of mixed responses were measured at the other institutions.[5]
- Significant shift from point-like to set-like.

"I think taking more data is better for their experiment." - Mixed

UR - Using Repeated Measurements

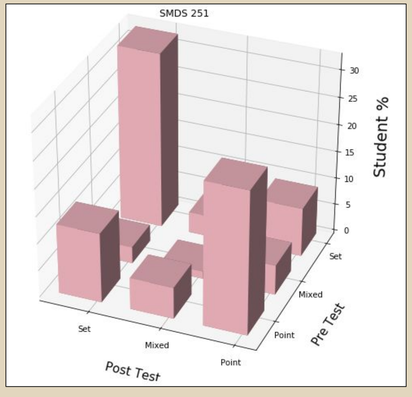


- Measures recognition to report an average or uncertainty for a data set.
- Overwhelming majority (91%) of students are set-like before and after instruction.
- Suggests that students already enter the course with expert reasoning in this context.

"I would write down the average for the experiment." - Set

"The average is useful because it accounts for deviation in the data." - Set

SMDS - Same Mean with Different Spread

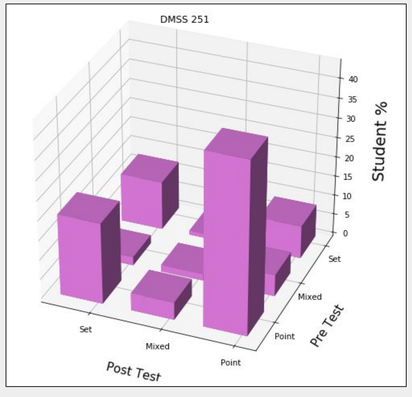


- Measures understanding that a data set with smaller spread is preferred.
- About 32% of students remain set-like and 26% remain point-like
- Suggests that students continue to hold similar levels of understanding about spread after instruction.

"Having a smaller uncertainty makes the average more accurate." - Set

"They got the same mean, that's all that matters." - Point

DMSS - Different Mean with Same Spread



- Measures knowledge of overlapping error between two datasets.
- Majority of students (44%) are point-like for pre test and post test.
- Moderate shifts from point-like to set-like

"The means are only off by 2mm, so they agree." - Point

"Both mean values are within the standard deviation of both datasets" - Set

References: [1] A. Buffler, S. Allie, and F. Lubben, Int J Sci. Educ. **23**, 37, (2001) [2] J. Cohen, Educ. and Psych. Measurement, **20**, 37, (1960) [3] B. Pollard, R. Hobbs, J. Stanley, D. Dounas-Frazer, and H. Lewandowski, 2017 PERC Proceeding, 312 (2018) [4] Codebook developed by R. Hobbs at CU [5] H. Lewandowski, R. Hobbs, J. Stanley, D. Dounas-Frazer, and B. Pollard, 2017 PERC Proceeding, 244 (2018)

Funding provided by the *Howard Hughes Medical Institute*.