Calibrating Multiple Graders

Reid Mumford, Physics & Astronomy, Lunch & Learn: 12/15/2017

General Physics Laboratory Courses

- Two-semester sequence
- 1-Credit courses
- Both labs taught every semester
- Multiple sections for each course:
  - 23 Sections: On-Semester (~500 students)
  - 6 Sections: Off-Semester (~100 students)
- Sections have up to 24 students each
- Sections taught & graded by Graduate TAs (~30 TAs)
- Teaching and grading styles vary between TAs.

Grading Overview

- Grades are based on best 9 of 10 lab activities.
- Activities consist of a Pre-lab Quiz and a Lab Note
- All activities are graded using the same rubric

Percentage of Points	Description
20%	Pre-lab Quiz
10%	Experiment Purpose
10%	Presentation and Analysis
10%	Results
10%	Uncertainty and Error Propagation
20%	Discussion
10%	Answers to Inline Questions
10%	Style

Example Rubric

- Students have direct access to the grading rubric
- “Lab Note Grading and Self-Assessment Rubric” PDF file

Results				
Poor – 1	**Fair – 2**	**Average – 3**	**Good – 4**	**Excellent – 5**
No attempt is made to present a final result.	A result is presented but it may be difficult to find, poorly formatted, and or not a result that is relevant to the lab.	A relevant result is presented but it may not be in standard form, may also be missing measurement units, and may be buried in text and difficult to find.	A correct result is given but may have a minor mistake in either labeling, significant figures, units, or standard formatting that prevents it from earning full points.	A result is presented. It is clear what is being measured. The result is easy to find, is presented in standard form with an associated uncertainty, the appropriate significant figures, and the appropriate measurement units.

Grading

Turnitin (TII)

- We adopted 3 years ago
- Famous for its plagiarism detection
- More importantly it provides efficient grading tools
- Integrates well with Blackboard (Bb)

Grading takes 2-4 hours each week

Graders are encouraged to:

- Be very demanding
- Leave lots of feedback

Examples: TII “Feedback Studio”

Lab 5: Ohm's and Kirchhoff's Laws

Experiment Purpose and Design

The purpose of the experiment is to test Ohm's law and Kirchhoff's laws by constructing series and parallel circuits and measuring the electric voltage drop across and the current through elements in the circuit. Elenco Snap Circuits were used to construct the circuits. Two digital multimeters were used, one as an ammeter to measure current and one as a voltmeter to measure the voltage drop from one point to another. A DC power supply was used as a battery. The voltmeter was set up to be parallel with the resistor(s) being measured. The ammeter was set up to be in series with the resistor(s) being measured.

Data Collection and Calculations

The voltage across a path, V, was measured in volts using a digital multimeter. The current through a path, I, was measured in milliamps using a digital multimeter. Ohm's law, $V = IR$ was changed into the more general form of $V = aI^b$. This general form was linearized into the equation $\ln(V) = \ln(a) + b\ln(I)$ by taking the natural log of both sides. Parameter a was calculated by taking the exponential of the y-intercept from the linearized equation $a = e^{\ln(a)}$. Parameter b and ln(a) were found using the LINEST function in Excel.

Presentation and Analysis

Table 1. Data for 5,100 Ω Resistor

Voltage (V)	Current (mA)	Current (A)	ln(I)	ln(V)	Error of ln(V)	Error of ln(I)
1.242	0.24	0.00024	-8.33	0.21	0.09	0.14
3.317	0.65	0.00065	-7.34	1.199	0.008	0.06
6.33	1.25	0.00125	-6.68	1.845	0.010	0.04
9.41	1.87	0.00187	-6.28	2.241	0.009	0.03
12.15	2.41	0.00241	-6.03	2.497	0.009	0.03
13.99	2.78	0.00278	-5.89	2.638	0.009	0.03
15.51	3.08	0.00308	-5.78	2.741	0.009	0.02

Power supply

Rubric

GPL Standard Lab Report

35 /40

Apply to Grade

Purpose 2
Average
3
1 5

Inline 1
Good
4
1 5

Data
Excellent
5

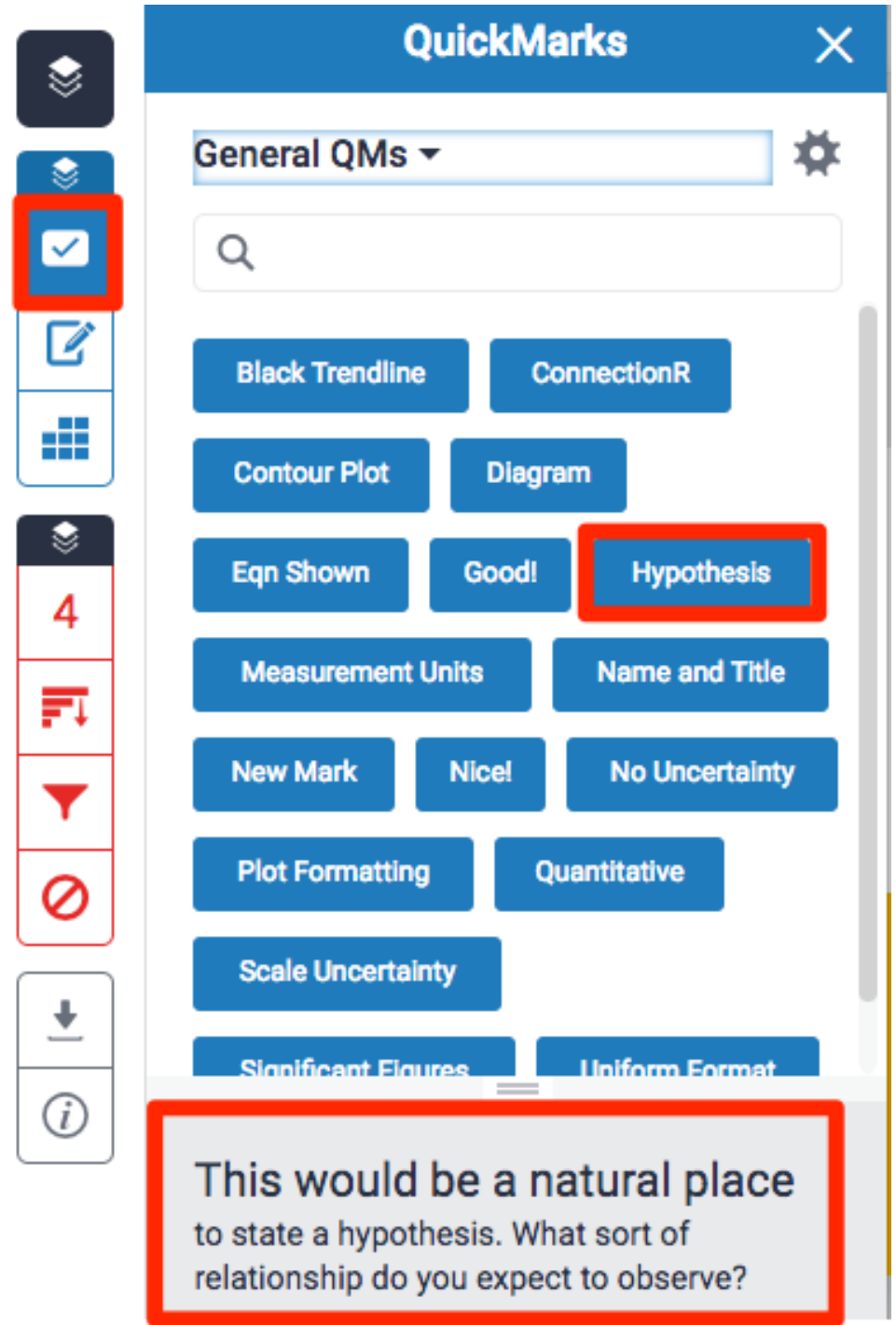
Purpose

In your own words, state the purpose of the experiment. In a couple of sentences, describe how you will use the available equipment to make your measurement. Include pictures or diagrams as appropriate.

Criteria	Scales				
	Poor	Fair	Average	Good	Excellent
Purpose 2 In your own words, state the purpose of the experiment. In a couple of sentences, describe how you will use the available equipment to make your	1.00 No attempt is made to define the purpose of the experiment OR describe the key points of the experimental design.	2.00 Confusing, inaccurate, or misleading description of how the equipment is used in the experiment. The purpose given is vague and misses the intention of the lab activity.	3.00 A description is attempted but may lack clarity and/or fails to identify all of the essential experimental details.	4.00 A clear and complete description is given but may include small mistakes or fail to identify an essential experimental detail. The description may also be too wordy.	5.00 The purpose of the experiment is clearly and accurately articulated. A brief and useful description of how the available equipment is used to make the measurement is given.
Inline 1 Include answers to specific inline questions as appropriate to the laboratory activity.	1.00 No attempt is made to address the questions.	2.00 An attempt is made but the response clearly demonstrates a lack of understanding.	3.00 The attempt addresses the questions but multiple mistakes are present. The attempt does not constitute a complete answer.	4.00 The attempt represents a sufficient answer for the specific question. A mistake or conceptual mis-understanding keep this response from receiving full points.	5.00 The attempt sufficiently meets the criteria required to completely and clearly answer the question.
Data List the quantities you measured directly in this experiment. For each quantity above, state the range of values over which you decided to make measurements.	1.00 No attempt is made to collect data.	2.00 An attempt is made at collecting data. The data do not make sense, contain obvious mistakes, and/or are clearly insufficient to observe trends in the physical system.	3.00 A reasonable attempt is made at collecting data. Too few data points were recorded. Multiple mistakes may exist. While measured antities may be identified, the	4.00 The measured quantities are identified. The data are collected in a way that adequately samples the range of interest. The experiment could be improved by taking more	5.00 The collected data are error-free. A large amount of data is recorded to fully sample the range of interest. The sampling technique is well justified. The data are sufficient to
Uncertainty Demonstrate error propagation calculations used to quantify the uncertainty of your measurement. Identify at least one source on random error in your experiment.	1.00 No attempt is made to approximate the uncertainty or identify sources of random and systematic uncertainty.	2.00 Uncertainty is occasionally assigned to observed/measured quantities. No error analysis is attempted. No attempt is made to identify sources of random and systematic	3.00 Uncertainty is consistently assigned to observed/measured quantities. An attempt is made at error analysis through the propagation of uncertainties and presented	4.00 Uncertainty is consistently (and correctly) assigned. Error propagation is clearly communicated and accurately calculated. Plausible sources of	5.00 Reasonable uncertainty estimates are assigned to all measured quantities. Error propagation formulas are clearly stated and used to accurately calculate uncertainty
Presentation Present the data you collected. Analyze your data using insightful	1.00 No attempt is made to present the recorded data in an organized way.	2.00 An attempt is made to present the recorded data but it is poorly	3.00 The recorded data are presented in an organized data table.	4.00 The data are presented in an organized and accurate data table.	5.00 The data are presented in an easy to read way. In all cases descriptive

Total Score: 35/40

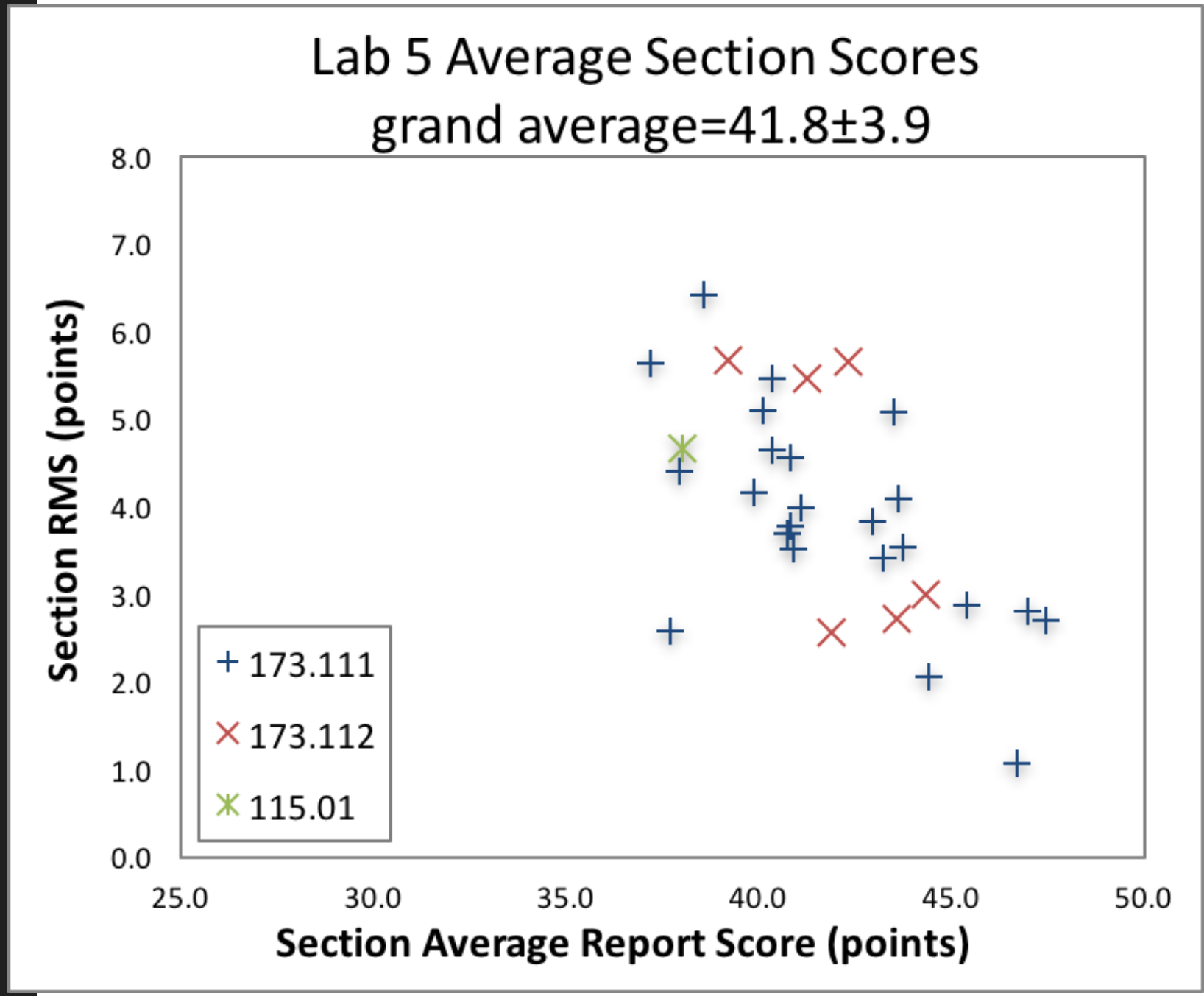
Apply to Grade | Close

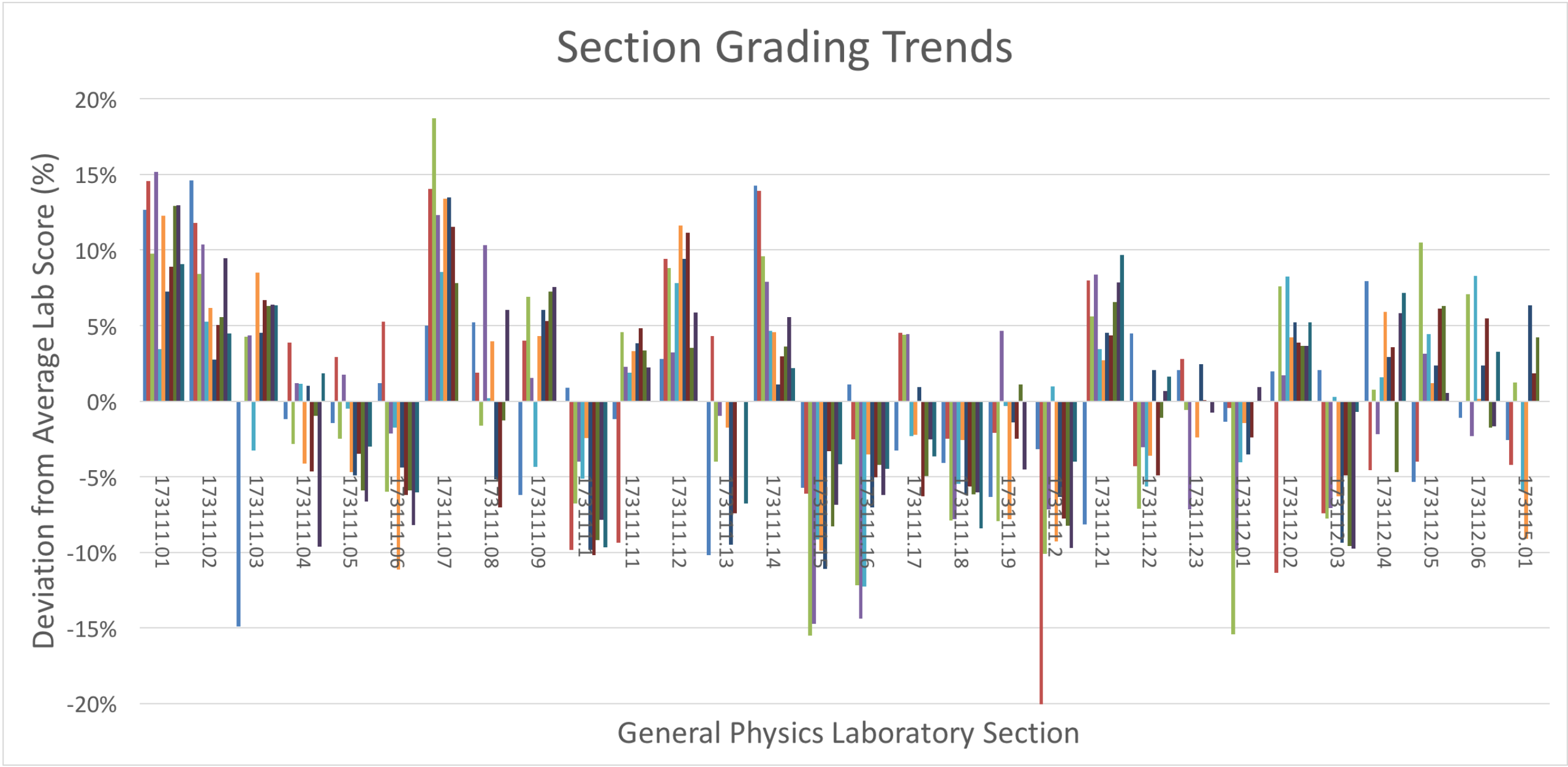

QuickMarks
General QMs
Black Trendline
ConnectionR
Contour Plot
Diagram
Eqn Shown
Good!
Hypothesis
Measurement Units
Name and Title
New Mark
Nice!
No Uncertainty
Plot Formatting
Quantitative
Scale Uncertainty
Significant Figures
Uniform Format
4
This would be a natural place
to state a hypothesis. What sort of
relationship do you expect to observe?

Calibrating Multiple Graders

Weekly Feedback

- Students are shown Average for their section
- Feedback to graders is vital
- Weekly Bb Grade Center downloads
  - Automated with browser plug-in: “Bulk URL Opener”
- Activity average and standard deviation are calculated
- Results are shared with the TAs




Section Grading Trends
Deviation from Average Lab Score (%)
20%
15%
10%
5%
0%
-5%
-10%
-15%
-20%
173111.01
173111.02
173111.03
173111.04
173111.05
173111.06
173111.07
173111.08
173111.09
173111.1
173111.11
173111.12
173111.13
173111.14
173111.15
173111.16
173111.17
173111.18
173111.19
173111.2
173111.21
173111.22
173111.23
173112.01
173112.02
173112.03
173112.04
173112.05
173112.06
173115.01
General Physics Laboratory Section

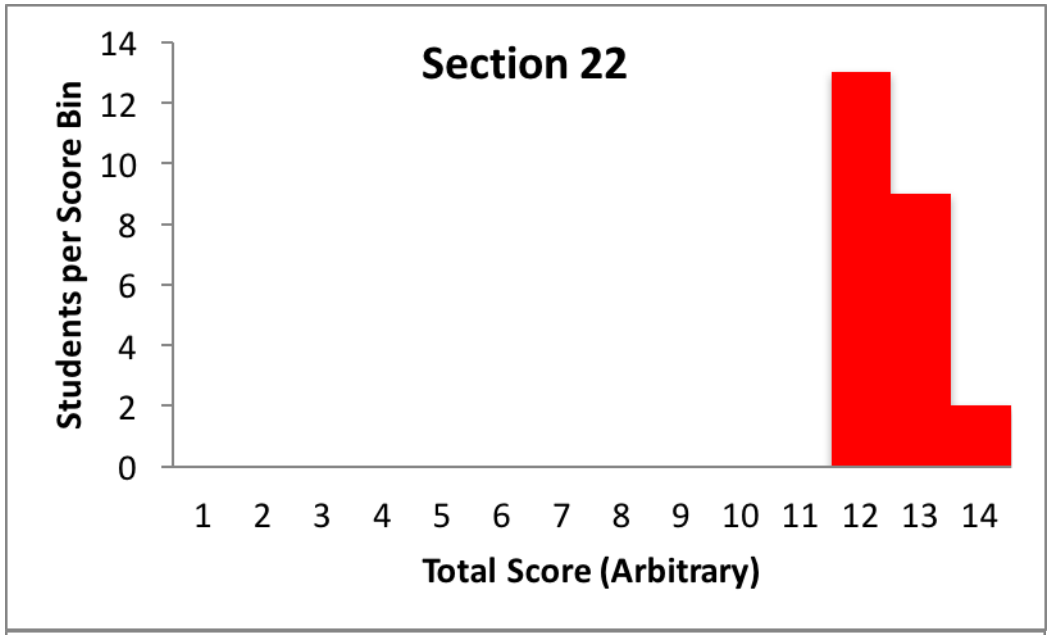
Final Grades

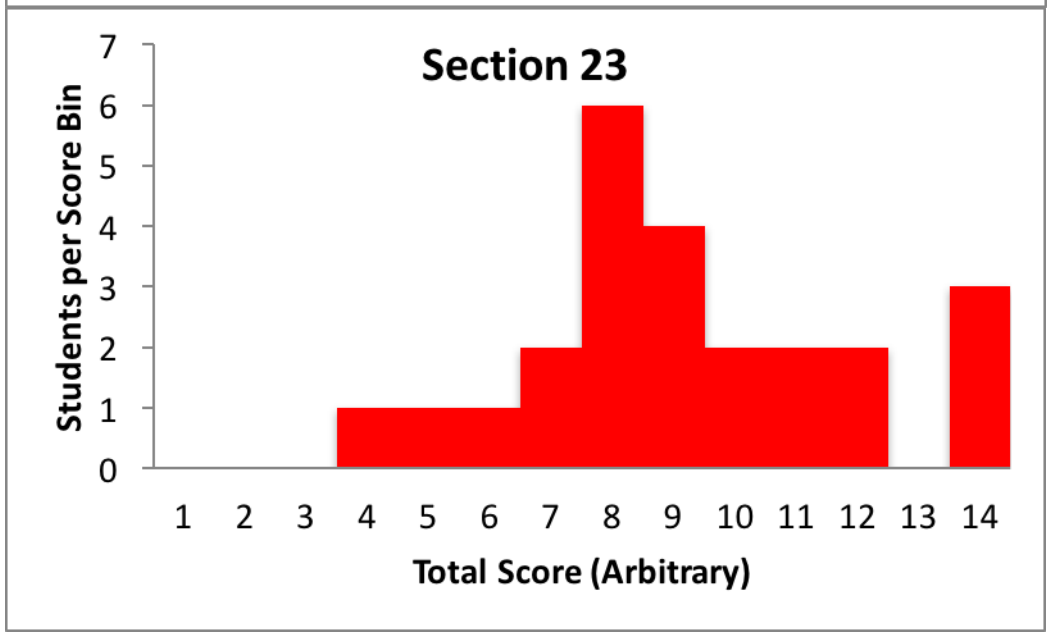
Final letter grades must be calculated

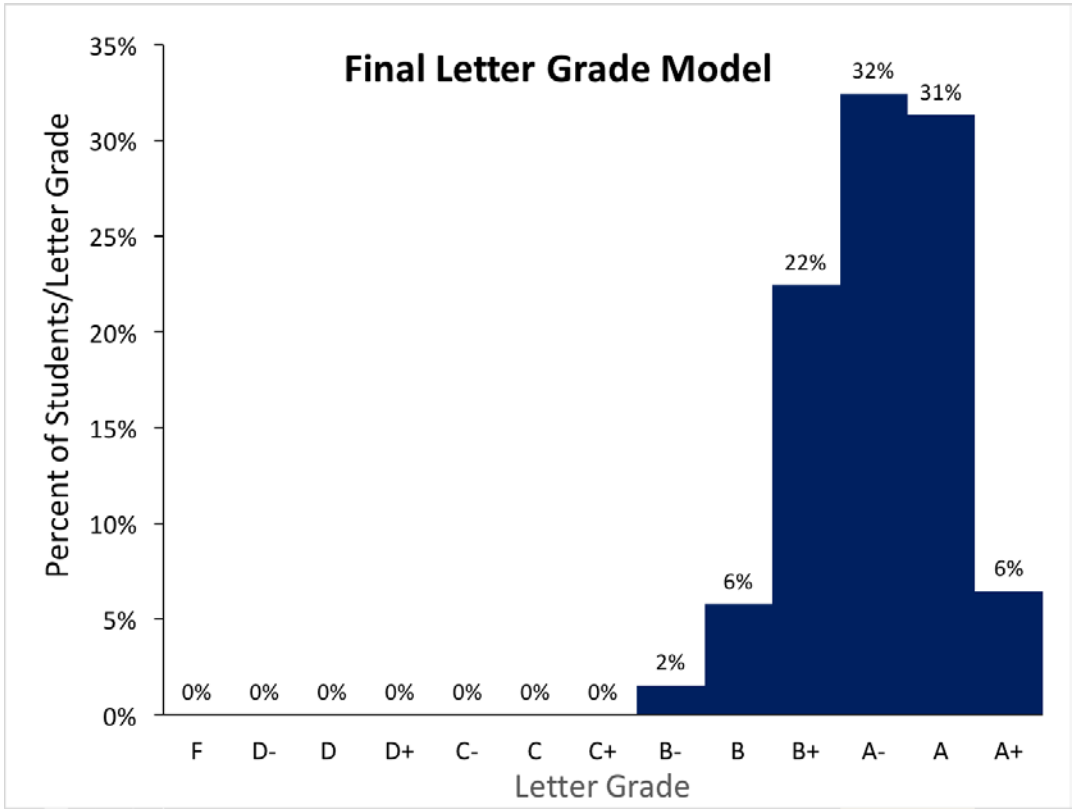
Should be no “easy” or “hard” sections of lab

Distribution should not vary (significantly) between sections

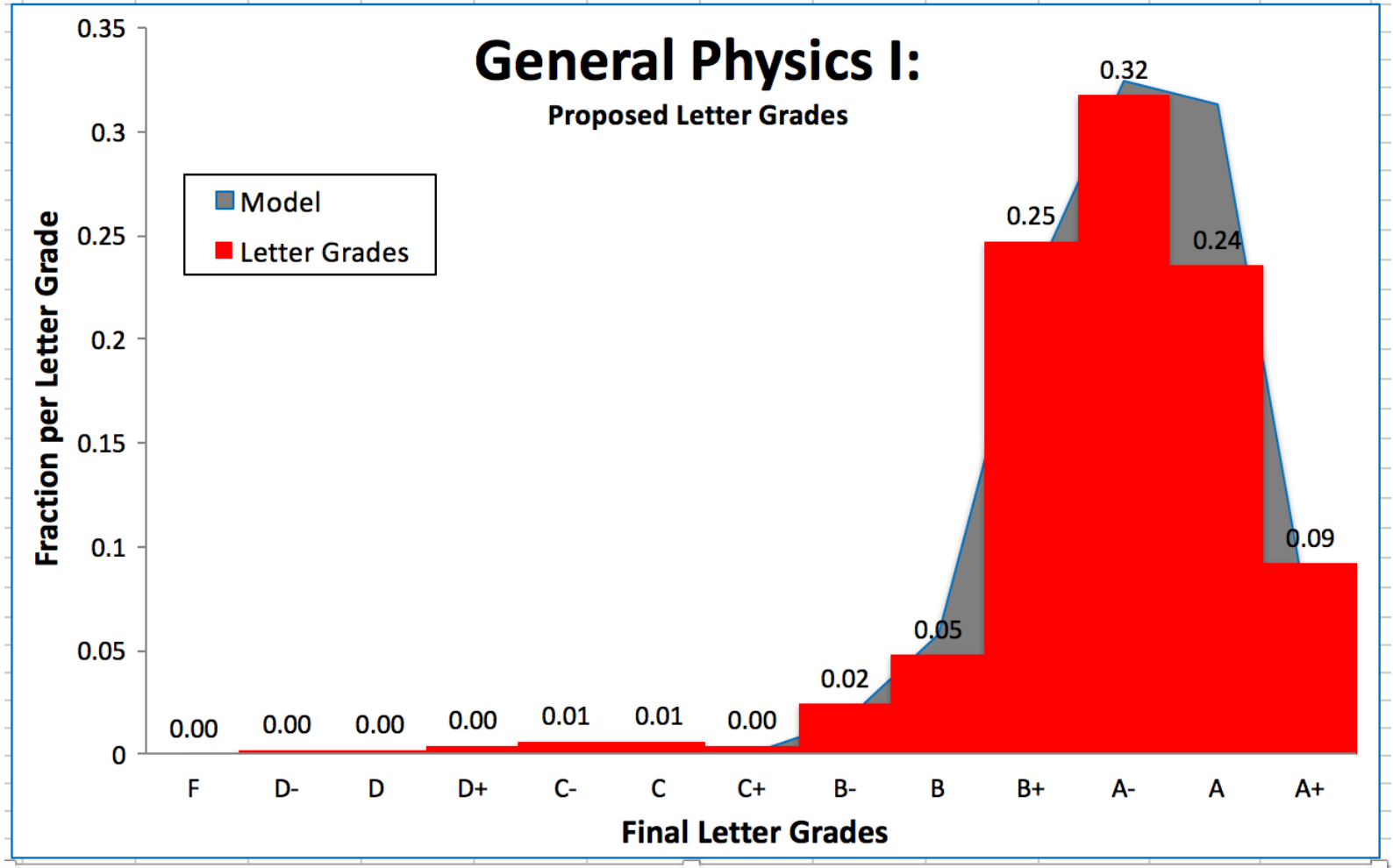
Calculating Grades: Per-section mapping

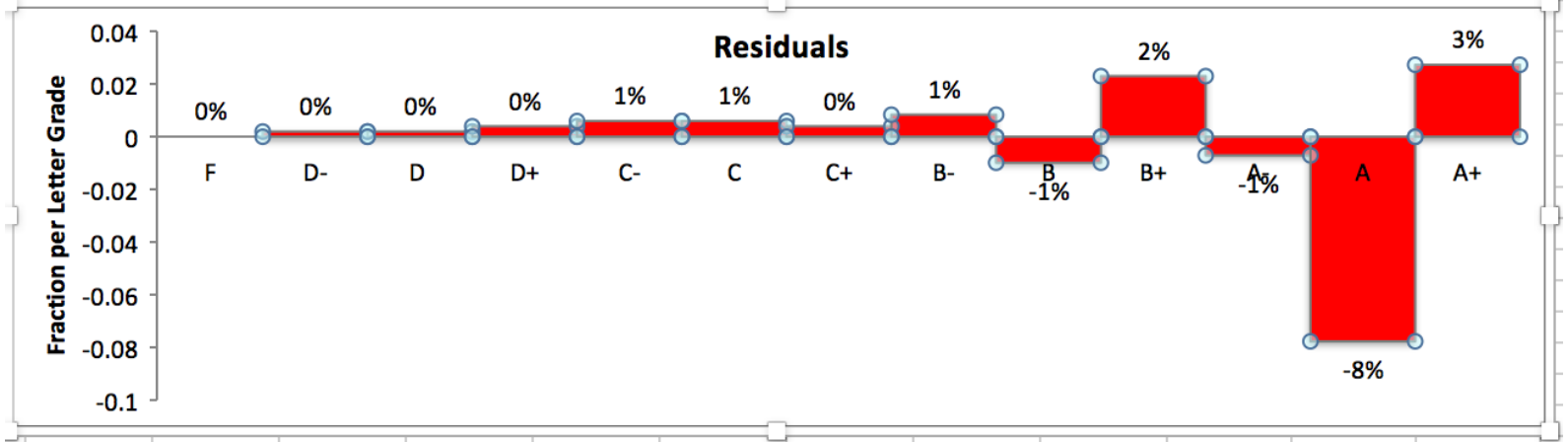






2-parameters to adjust distributions:
Average and Standard Deviation


General Physics I:
Proposed Letter Grades
Model
Letter Grades
Fraction per Letter Grade
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.00
0.00
0.00
0.00
0.01
0.01
0.00
0.02
0.05
0.25
0.32
0.24
0.09
F
D-
D
D+
C-
C
C+
B-
B
B+
A-
A
A+
Final Letter Grades


Residuals
Fraction per Letter Grade
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
0%
0%
0%
0%
1%
1%
0%
1%
-1%
2%
-1%
-8%
3%
F
D-
D
D+
C-
C
C+
B-
B
B+
A-
A
A+

Take-Aways

Calibrating Multiple Graders is not easy

Need tools to handle multiple sections efficiently

Rubrics help but do not solve the calibration problem

Regular feedback to graders is essential

Limit of our system: Student standing is ambiguous

Future plans:

- Give students better understanding of course standing
- Calculate a per-section curve each week
- Obstacles: Technical issues and larger time investment required