# Reforms and measurements in introductory physics

Marcos Caballero[1] Michael Schatz[1] Keith Bujak[2]
Richard Catrambone[2] M. Jackson Marr[2] Matthew Kohlmyer[3]

[1]School of Physics, Georgia Institute of Technology
[2]School of Psychology, Georgia Institute of Technology
[3]Department of Physics, North Carolina State University

April 30, 2010

Supported by the NSF (DUE-0942076, DUE-0618519)
Some work in collaboration with
North Carolina State University (DUE-0618504) and Purdue University (DUE-0618647).

# Outline

# Introductory Physics @ Georgia Tech

## Two Semester Sequence

- Semester 1 – Mechanics
- Semester 2 – Electromagnetism

## Boundary Conditions for Intro Physics

- $\sim$ 1600 students per semester
- 83% engineering, 17% science majors
- 3 hours of Lecture (150-250 students)
- 3 hour Lab/Recitation (25-40 students)

# Shortcomings of Traditional Curriculum (TRAD)

## Content unchanged for decades

- 19th century (or earlier) concepts
- Focus on analytic solutions of special cases

## Difficulties at Tech

- GPA lower than other intro courses
- High D/F/W rate (as high as 25%)
- Unpopular with students
- External review criticized structure, outcomes

# Outline

# Reform Curriculum: Matter and Interactions (M&I)

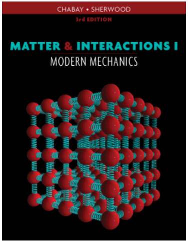
R. Chabay & B. Sherwood, Wiley, 2010
www.matterandinteractions.org

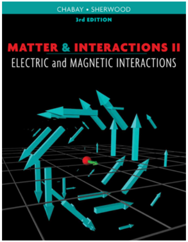
## Modern content

- Fundamental principles
- Atoms and structure of matter
- Relativity and quantum physics
- Macro/micro connections

## Modern tools/techniques

- Computer modeling



# Implementation @ GT

## Infrastructure Preparation

- Local expert: Hire post-doc (1/06-9/08)
- Train teaching assistants (Spring 06, on-going)
- Laboratory equipment purchase/construction (Spring 06, Fall 07, Spring 10)

## Faculty Preparation: Apprenticeship Model

- Junior faculty
  - Pair with experienced instructor
  - Provide logistical support
- Senior faculty
  - Same plus financial incentive

# Gradual Implementation

## Student Enrollment and Faculty Adoption

| Semester | M&I Semester 1 | M&I Semester 2 | Faculty w/M&I experience |
|---|---|---|---|
| Summer 06 | 40 students | None | 0 |
| Fall 06 | 120 students | 45 students | 1 |
| Spring 07 | 200 students | 150 students | 2 |
| Summer 07 | None | 150 students | 3 |
| Fall 07 | 150 students | 300 students | 4 |
| Spring 08 | 300 students | 300 students | 4 |
| Summer 08 | 150 students | 150 students | 4 |
| Fall 08 | 300 students | 450 students | 6 |
| Spring 08 | 500 students | 300 students | 6 |
| Summer 09 | 250 students | None | 6 |
| Fall 09 | 400 students | 550 students | 7 |

# Is reform doing any good?

## Compare student performance

- Traditional (control) vs M&I (reform)

## In class measurements

- Concept inventories
- Common final exam problems
- Performance in follow-on courses

## Out of class measurements

- Think aloud protocol studies

# Outline

# Concept Inventory – Semester 2 (E&M)

## Brief E&M Assessment (BEMA)

- 31 item multiple choice test covers all E&M
- Qualitative and short quantitative questions
- Items common to M&I and TRAD course
- Administer "pre-test" at beginning of course, "post-test" at end, measure gains

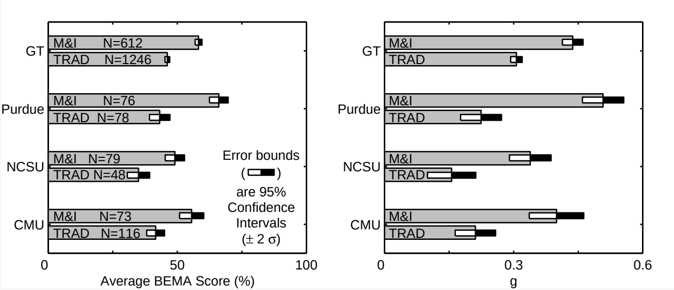
## Topics

- Electrostatics (ES)
- DC Circuits (DC)
- Magnetostatics (MS)
- Faraday's Law and Induction (FL)

# Performance at 4 Institutions

*Phys. Rev. ST Phys. Educ. Res.* 5, 020105 (2009)

## M&I outperforms TRAD at all Institutions on the BEMA (E&M)

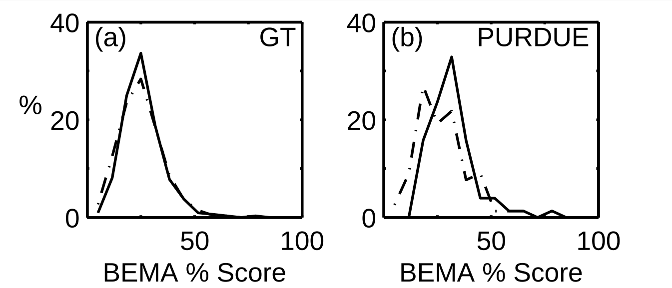



## Computing $G$ and $g$

- Raw Gain, $G = \text{Post\%} - \text{Pre\%}$
- Normalized Gain, $g = G/(100\% - \text{Pre\%})$

# Pre-tests at GT and Purdue

## Incoming BEMA scores similar



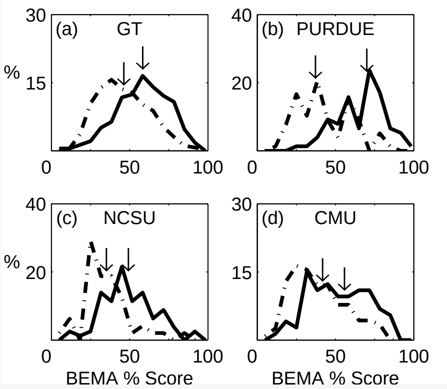

- Distributions similar (Wilcoxon test, p $\sim$ 0.30)

## Essential demographics similar

- Grade in Physics I, Calculus
- GPAs, SAT scores

# Post-test at all Institutions

## Outgoing BEMA scores favor M&I



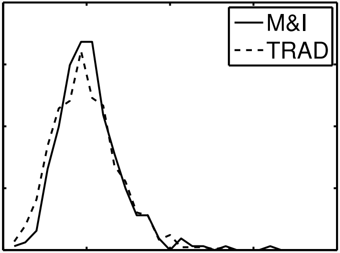

- Higher means (Wilcoxon test, p $<<$ 0.001)

# Focus on Large N – Georgia Tech

## Pre and Post-test scores

Pre-test

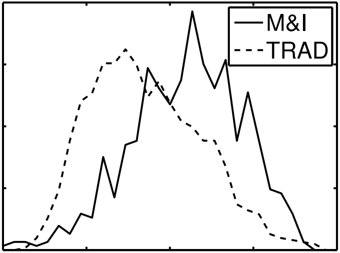
$\bar{X}_{\mathrm{MI}} = 25.9\%,\ \bar{X}_{\mathrm{TRAD}} = 24.8\%$




$N_{\mathrm{MI}} = 321,\ N_{\mathrm{TRAD}} = 1319$

Post-test

$\bar{X}_{\mathrm{MI}} = 58.2\%,\ \bar{X}_{\mathrm{TRAD}} = 46.1\%$



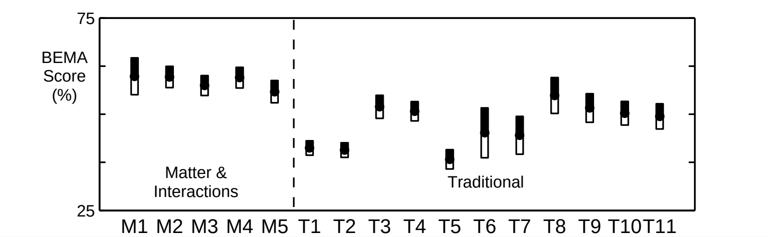

$N_{\mathrm{MI}} = 612,\ N_{\mathrm{TRAD}} = 1246$

- Jagged distributions – Discrete scores

# Post-test BEMA Scores by Section

## Large variation for TRAD (pedagogy, instructor)

$$\bar{X}_{\mathrm{MI}} = 58.2\% \quad \bar{X}_{\mathrm{TRAD}} = 46.1\%$$



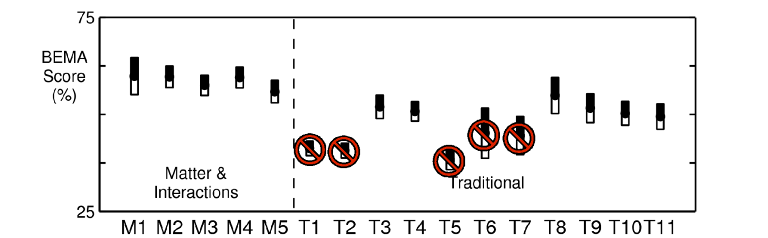

- No "instructor effect" for M&I (Kruskal-Wallis test)
- Significant "instructor effect" for TRAD (Kruskal-Wallis test)

# Post-test BEMA Scores by Section

## Consider TRAD Instructors using Active Engagement

- Sections T3,T4,T8,T9,T10,T11

## Instructors that use "clickers"




- TRAD instructors using "clickers", $\bar{X}_{\mathrm{TRAD}} = 51.3\%$
- Lower than M&I instructors, $\bar{X}_{\mathrm{MI}} = 58.2\%$ (Wilcoxon test)

# Performing an Item Analysis

## Compare Performance per Question

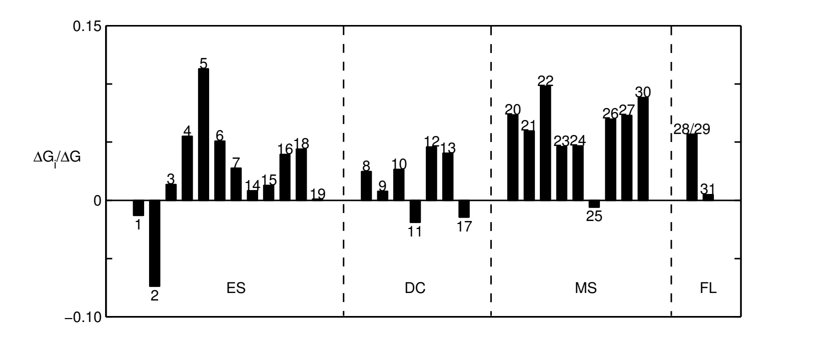
- Performance is gauged by raw gain
  $G = \mathrm{Post}\% - \mathrm{Pre}\%$
- Questions can be grouped by topic

## Computing Fractional Differences

- Overall difference, $\Delta G = G_{\mathrm{MI}} - G_{\mathrm{TR}}$
- Item difference, $\Delta G_i = G_{i,\mathrm{MI}} - G_{i,\mathrm{TR}}$
- Fractional difference, $\Delta G_i / \Delta G$

# Difference in Performance per Question

## Fractional Difference identifies Strong Contributions to Performance




- Electrostatics (ES) 27.8%
- Magnetostatics (MS) 54.9%
- DC Circuits (DC) 11.8%
- Faraday's Law (FL) 6.2%

# Concept Inventory – Semester 1 (Mechanics)

## Force Concept Inventory (FCI)

- 30 item multiple choice test covering force and motion
- Qualitative questions, emphasizes constant force motion
- Designed in context of a TRAD curriculum
- Administer “pre-test” at beginning of course, “post-test” at end, measure gains

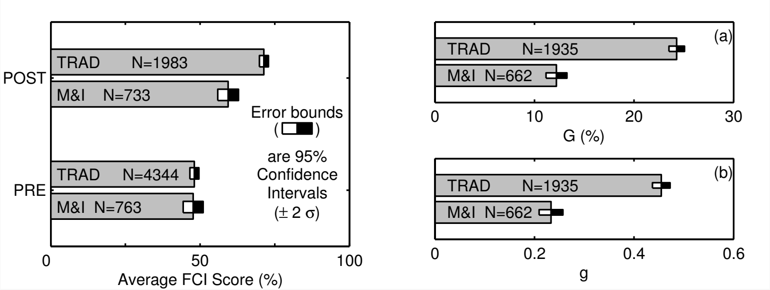
## Topics

- Kinematics
- Newton’s Laws (1st Law, 2nd Law, & 3rd Law)
- Force Identification

# Performance at Georgia Tech

*Am. J. Phys.* (2010, in preparation)

## TRAD outperforms M&I on the FCI: $\bar{X}_{\mathrm{TRAD}} = 71.3\%$, $\bar{X}_{\mathrm{MI}} = 59.3\%$

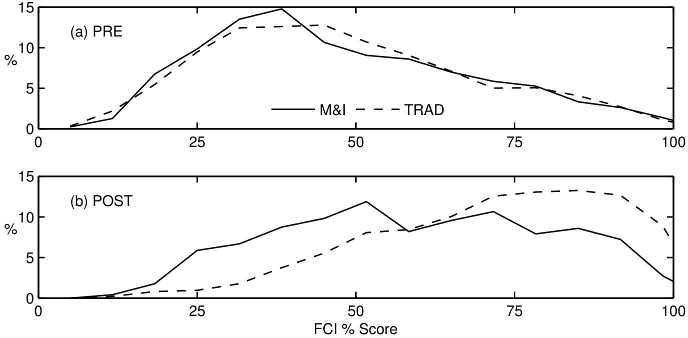



## Other measures not statistically different

- GPAs, SATs, Incoming FCI, etc.
- Pedagogy (interactivity, presentation, etc.) very similar

# Distributions of FCI scores

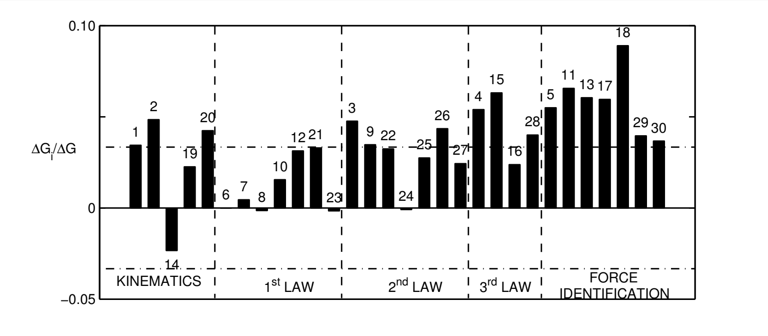
## Pre and Post-test scores




- Data has been binned to reduce jagged appearance

# Difference in Performance per Question

## Fractional Difference illustrates where M&I under-performs




- Experts' *a priori* categorization
- Fraction a question contributes to overall difference (TRAD-M&I)

# Underlying patterns to responses

## Principal Component Analysis

- Mathematically sound
- Do clusters mean anything?

## Factor Analysis

- Orthogonal vs. Non-orthogonal rotations?
- Controversy (Heller, 1995 & Hestenes 1995)
- “Best” method for binary measures?

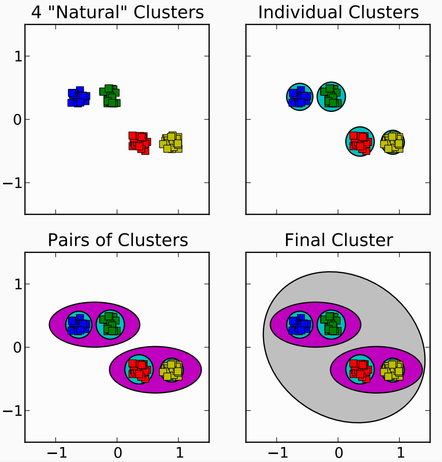
## Cluster Analysis

- Mathematically sound
- Distance measures clearly defined
- Consistent results with different measures

# Cluster Analysis

## Finding Patterns in Data

- Hierarchical Cluster Analysis



# Cluster Analysis of the FCI

## Binary Data Metric

- Hamming Distance

$$D_h = \frac{C_{tf} + C_{tf}}{N}$$

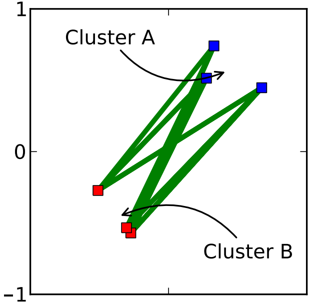
- ex. [1001] and [1010], $D_h = 2$

| | | | |
|---|---|---|---|
| [1001] | [1010] | $\rightarrow$ | [1010] |
| [1001] | | | [1011] |
| [1001] | [1001] | $\leftarrow$ | [1001] |

## Linkage Criteria

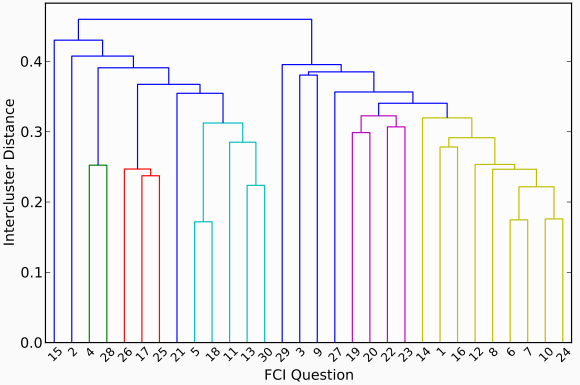
- Mean linkage clustering

$$\bar{d} = \frac{1}{|A||B|} \sum_{a \epsilon A} \sum_{b \epsilon B} D_h(a, b)$$



# Cluster Analysis of M&I data

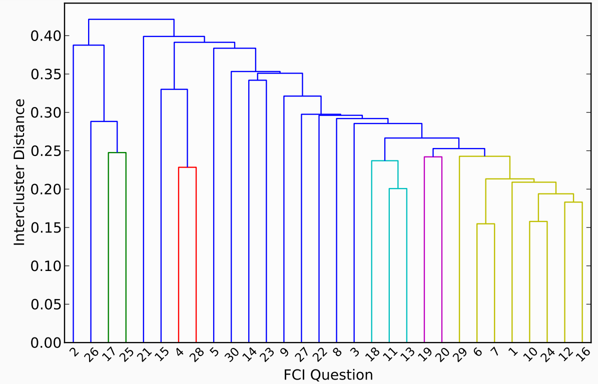
## Agglomerative categorization of problems (M&I)




- Students' pattern of performance defines clusters
- Cluster [(5, 18), 11, (13, 30)] similar to *a priori* grouping

# Cluster Analysis of TRAD data

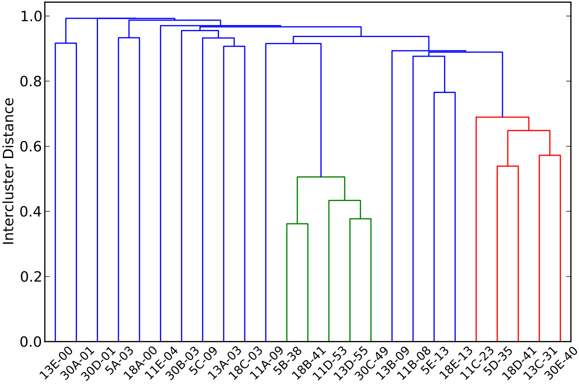
## Agglomerative categorization of problems (TRAD)




- Clusters are similar to M&I clusters
- Cluster [(11, 13), 18] similar to *a priori* grouping

# Pattern of Responses

## [(5, 18), 11, (13, 30)] Response Cluster (M&I)




- Jaccard Distance

$$D_j = \frac{C_{tf} + C_{tf}}{C_{tf} + C_{tf} + C_{tt}}$$

- Responses equally weighted (no $C_{ff}$)
- Include fraction of students

- 2 dominant responses
  - Correct (green) and "Force in direction of motion" (red)

# Common Final Exam Questions

## Similar (Dismal) Performance

- Percentage of “Mostly Correct” Responses

| Question | TRAD. | M&I |
|---|---|---|
| Mech. 1 | 13% | 26% |
| Mech. 2 | 29% | 24% |
| Mech. 3 | 7% | 8% |
| Mech. 4 | 21% | 17% |
| Mech. 5 | 59% | 49% |
| E&M 1 | 10% | 17% |
| E&M 2 | 22% | 39% |
| E&M 3 | 20% | 29% |

# Average GPA in Advanced Courses

## Similar Performance

- COE 2001(Engineering Statics) (Requires Mechanics Prerequisite)
    - TRAD: 2.79 ($\pm$ 0.05) (N = 1695)
    - M&I: 2.81 ($\pm$ 0.11) (N = 359)
- ECE 3025(Electromagnetics) (Requires E&M Prerequisite)
    - TRAD: 2.93 ($\pm$ 0.15) (N = 144)
    - M&I: 2.95 ($\pm$ 0.28) (N = 24)

## No significant difference

- Wilcoxon test, COE 2001 ($p \sim 0.30$)
- Wilcoxon test, ECE 3025 ($p \sim 0.40$)

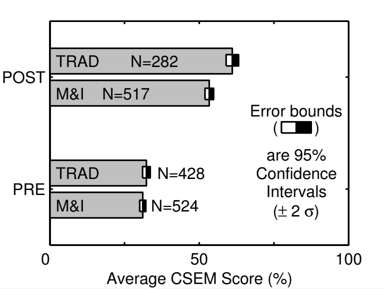
# Lessons Learned

## More and Better Measurements are Needed

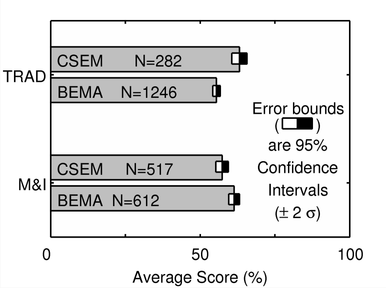
- Measurements are difficult
- Beware of over-reliance on particular measurements

## Another E&M Concept Inventory: CSEM

CSEM scores (Fall 2009)




BEMA overlap (7 questions)



# Concept Inventories

## Aren't BEMA & CSEM Equivalent?

- Yes, (for same (Traditional) curriculum)
  S. Pollock, *PERC Proceedings* (2008)

## What about the FCI?

- Couched in TRAD language & examples
- M&I performance unaffected by prompting
  Caballero, unpublished (2008)
- FMCE? Similar differences in performance

## Inventories equivalent for different curricula?

- More (and different) measurements needed
- Need variety

# What are we doing now?

## Characterizing the observed FCI differences

- Homework categorization, continuing cluster analysis

## Exploring students' mechanics knowledge

- FCI & Core Mechanics Think Aloud studies

## Developing computational knowledge

- Computer modeling homework
- Evaluation of modeling skills

## Diversifying assessment

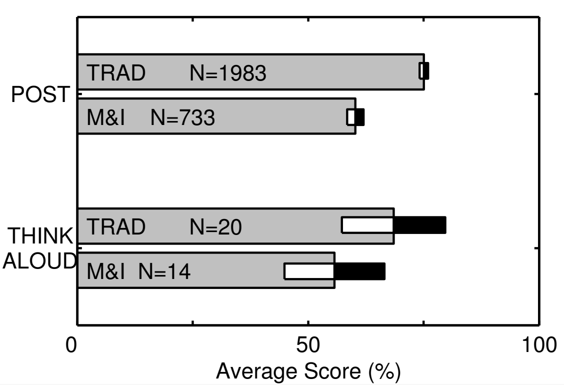
- Qualitative - MIT Survey of Mechanics
- Attitudinal - CLASS, GT-designed "Attitudes about Modeling"

# Outline

# Think Aloud Study

## Semester 1 Concept Inventory

- Audio and video record subjects solving subset of 10 FCI questions
  - Subset had high contributions to the difference in scores



# An Example from the Think Aloud Study

## FCI Question #1

Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant of time. The time it takes the balls to reach the ground below will be:
(A) about half as long for the heavier ball as for the lighter one.
(B) about half as long for the lighter ball as for the heavier one.
(C) about the same for both balls.
(D) considerably less for the heavier ball, but not necessarily half as long.
(E) considerably less for the lighter ball, but not necessarily half as long.

- Correct Response (C) - 90% TRAD, 57% M&I
- Major distractors (A & D) - 10% TRAD, 36% M&I

| **Solution** | **TRAD (%)** | **M&I (%)** |
|---|---|---|
| Determined acceleration was constant | 0 | 36 |
| "mass doesn't matter" | 60* | 21 |
| use of kinematics equations | 40 | 0 |
| recall from previous exercise | 20 | 0 |

*Half of these students also used kinematic equations

# Suggestive Results from Transcript Analysis

## Lessons from FCI transcripts

- M&I students fail to employ $\Delta\vec{p} = \vec{F}_{net}\Delta t$ (NO mention of momentum at all)
- M&I students confuse components of $\vec{F}_{net}$ and agent forces
- Many students revert to naive/incorrect notions
- Some recall (often, incorrect/incomplete) memory of HS physics
- Students select correct responses without a deep understanding

## Core Mechanics problems

- Developed and implemented new problems (non-constant forces)
- Think aloud work on-going (transcription almost complete)

# Outline

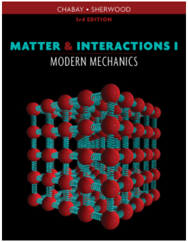
# The Matter and Interactions Mechanics Course

## M&I differs from a "typical" physics course

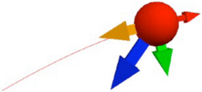
- Emphasizes a principles based approach
  - $\Delta\vec{p} = \vec{F}\Delta t$
  - $\Delta E = W + Q$
  - $\Delta\vec{L} = \vec{\tau}\Delta t$

- Introduces the ball and spring model of matter
  - Young's modulus, Speed of sound
  - Statistical Mechanics, Temperature

- Uses modern tools (computer simulation)
  - Iterative view of motion (Non-constant forces)
  - Computer modeling laboratories




Student Model
Projectile with Drag

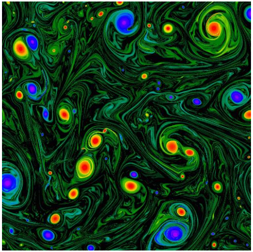


# Why Computer Modeling Homework?

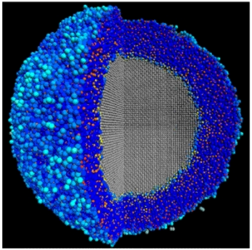
## Why Computer Modeling?

- Third pillar of science and engineering
  - Theory, Experiment, Computation
- Explore "intractable" systems
  - Effects of air resistance
  - 3D spring with viscous drag
- Simulate "impossible" experiments
  - Elliptical orbit
  - 3 body problem
- Visualize the problem
  - Observing the motion, physical vectors
  - Plotting of energy-time series

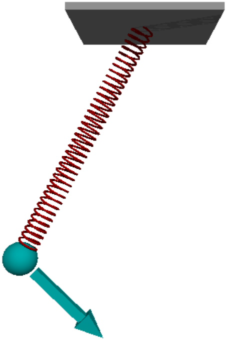
Source: UCLA Geophysics



Source: USC Advanced Computing



# Why Computer Modeling Homework?



3D Spring Model
cyan arrow represents $\vec{p}$

## On Homework Assignments?

- Large service course
  - Online homework system; no hand graded homework
  - Randomization does not deter "short-cuts" (MIT - Palazzo, et.al.)
  - Closed form solutions (Google, Yahoo! Answers, Wolfram|Alpha, etc.)
- Some never write programs
  - Programming "person" in lab group
  - Internal cost-benefit calculation
- Not a novelty
  - Another tool for solving problems
  - Visualization might help intuition

# Method of Implementation

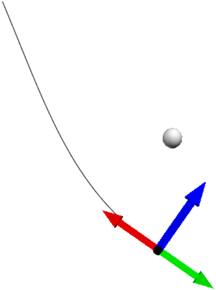
## Based on Computer Modeling Labs

- Numerical integration and differentiation
- Must be solved numerically

## Implementation facilitated by WebAssign

- Generate and store $\sim$ 400–800 realizations
- Randomize realization per student
- Receive 2 realizations
  - test case: solutions given
  - graded case: no solutions

- Numeric Questions
- Visualization Questions



Satellite trajectory
Arrows:
Red - $\vec{F}_{\parallel}$
Blue - $\vec{F}_{\perp}$
Green - $\vec{p}$

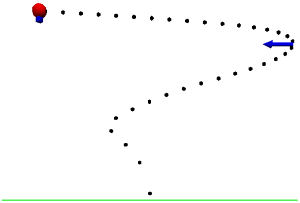
# A Typical Week

## Lab Program

- Position data file $\rightarrow \vec{F}_{\mathrm{net}}$
- Integrate, determine $\vec{r}_f$, $\vec{v}_f$
- Visualize $\vec{p}$, $\Delta\vec{p}$, and $\vec{F}_{\mathrm{net}}$

## Homework Assignment

- Reproduce work done in lab (new data file)
- Compute and visualize $\vec{F}_{\mathrm{net},\parallel}$ and $\vec{F}_{\mathrm{net},\perp}$
- Visualize $\vec{p}_f$, $\Delta\vec{p}$, $\vec{F}_{\mathrm{net}}$
- Graded for correctness, not completion



Weather Balloon Trajectory
blue arrow represents $\vec{F}_{\mathrm{net}}$

# So...how did it go?

## Instructor Perspective

- Relatively straight-forward implementation
- Minor hiccups caught early on:
  - Installation issues
  - Test cases
- Confirmed suspicions; Some students never write programs

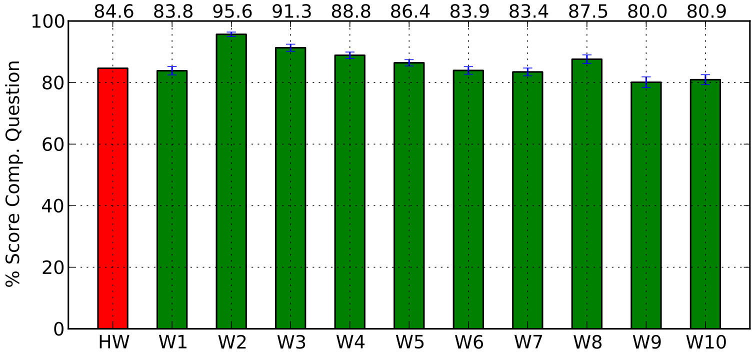
## Student Perspective

- Questions not treated as "special"
  - Questions tacked on to lab assignment
  - Questions appear on homework assignment each Monday
  - Similar weight as a single homework question
- Anecdotal evidence that students start homework earlier

# So...how did they do?

## Scores similar to average homework question

Homework Avg. (n=238): 84.60 %  Comp. Questions (n=10): 85.86 %



# Computer modeling in a controlled environment

## 4th Hour Exam Extra Credit

- 15 minute (password-protected) extra credit problem
  - integrate a central force
  - randomized conditions, force law, syntax
  - compute $\vec{r}_f$ and $\vec{v}_f$

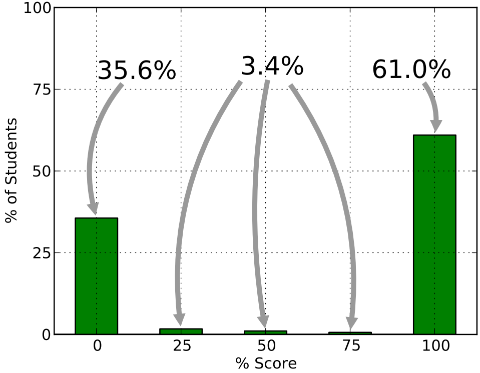
## Challenges

- Logistical problems
  - tabbing issue, feedback
  - time constraint
- Physics/modeling problems
  - error in sign of force, overflow error
  - adding vectors & scalars, etc.
- Instructors struggled with implementation

# Computer modeling in a controlled environment

## Proctored Lab Assignment

- Logistical changes
- Feedback from test case
- Lowered time constraint (25 min)
- No TA help ("Read carefully...")



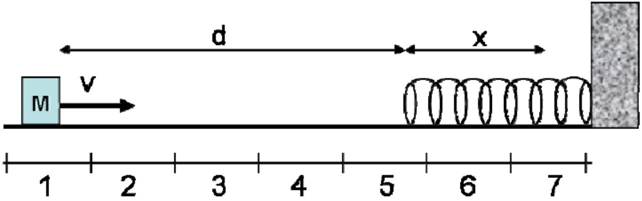

- Mean: $62.42 \pm 0.32$ %
- N = 469
- Most common "physics" mistake: sign error
- Other syntax errors: vector + scalar, etc.
- Analysis continues...

# Outline

# MIT Mechanics Assessment

## Another Concept Inventory?

- Problems are more complex
- Cover most of mechanics
- Process and reasoning driven




- Validation and reliability testing (Summer 2010)

# Attitudinal Surveys

## Colorado Learning Attitudes about Science Survey

- No attitudinal data from GT
- Similar results to CLASS study?
  Adams, et.al., *Phys. Rev. ST Phys. Educ. Res.* 2, 010101 (2006)
- Curricular differences?

## GT designed Attitudes/Impressions of Computer Modeling

- Similar design to CLASS
- Validating survey (i.e., colleague & student review)
- Curricular differences?
- Correlation of beliefs with student performance

# Outline

# Collaborators and Friends

## Thanks to:

- Ed Greco, Daniel Borrero, Huseyin Kurtuldu, Juan-Jose Lietor-Santos, & Adam Perkins (GT Physics)
  – Implementation, Problem Testing, Sanity
- Ruth Chabay & Bruce Sherwood (NCSU & CMU Physics)
  – Discussions, BEMA
- Mark Haugan, Deborah Bennett, Lynn Bryan, Dan Able, & Melissa Yale (Purdue Physics) – BEMA
- Lin Ding (OSU Physics) – BEMA

## Additional thanks:

- Marty Jarrio, Eric Murray, Andrew Scherbakov (GT Physics), & Bob Hume (GT OMED)
  – Discussions, FCI/BEMA/Exams

# Closing Remarks

## Future Work (Fall 2010 and beyond)

- Computer modeling skill development and measurement
- Attitudinal inventory development, measurements, & comparisons
- Full description (disclosure) of concept inventory measurements
- Comprehensive measurement of conceptual understanding using think aloud protocols
- Novel problem study based on Kohlmyer's work
- Intervention (Precision Teaching) and measurement

## More info?

- GT PER Group - *www.physics.gatech.edu/gtper*
- Contact me - *caballero@gatech.edu*