Patterns of Behavior in Online Homework

Colin Fredericks

The Point

- Everyone assigns homework.
- Very few studies done on…
  - ***How*** students do homework,
  - How they do it ***best***, and
  - The specific ***benefits*** of doing it a particular way.
- How much can we learn from homework data?

The Other Point: **Analysis**

- Part of *Researching the Role of Analysis*
- Analysis of problems is a powerful tool
  - Enhances physical intuition
  - Speeds solution
- Does specifically practicing analysis help, or does it only come with time?[1]

1) E. Kim and S-J Pak, *Students do not overcome conceptual difficulties after solving 1000 traditional problems*, Am. J. Phys. 70 (7), July 2002

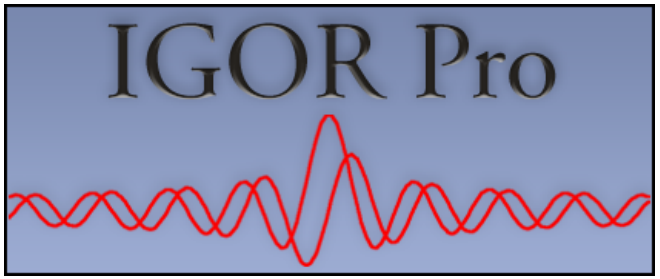
Data Sources

- Electronic homework at UMass Amherst
- Performance data such as exams and final course grades
- Surveys

- Physics 151, Fall 2003
  - 250 students
  - Eng/Chem/CS
  - 140,000 rows of data
- Physics 181, Fall 2005
  - 55 students
  - Physics/Astro
  - 8,500 rows of data

Methods and Tools

- Excel and Igor
- Correlation Factors
- Principal Component Analysis



Correl PCA Next

Correlation Factors

- Shows degree of linear relationship between two variables
- r^2 estimates amount of variance accounted for by a particular variable

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

PCA Next

What is PCA?

- Part multilinear modeling, part data reduction scheme
- Returns orthogonal vectors that are linear combinations of the original data
- Used in wide variety of fields: chemistry, social science, marketing

- Can be used to group items
- Can be used to identify random data, sort of

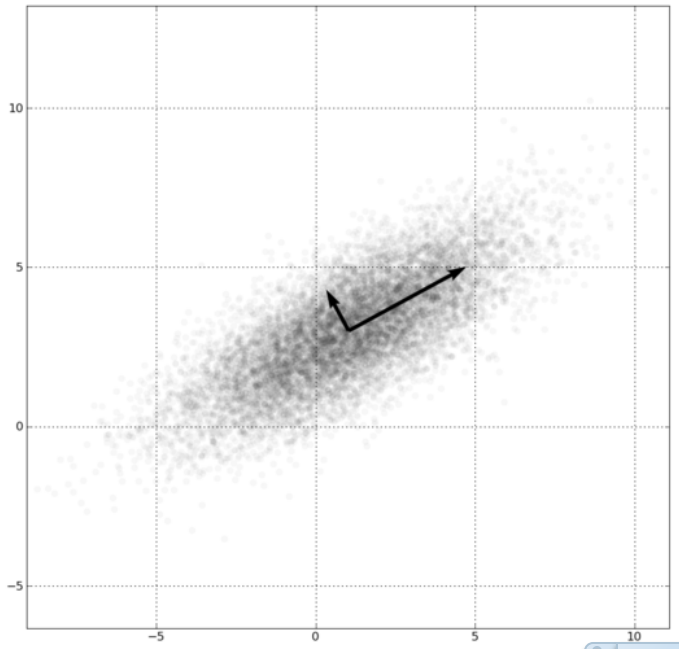


Image from https://en.wikipedia.org/wiki/File:GaussianScatterPCA.png

Next

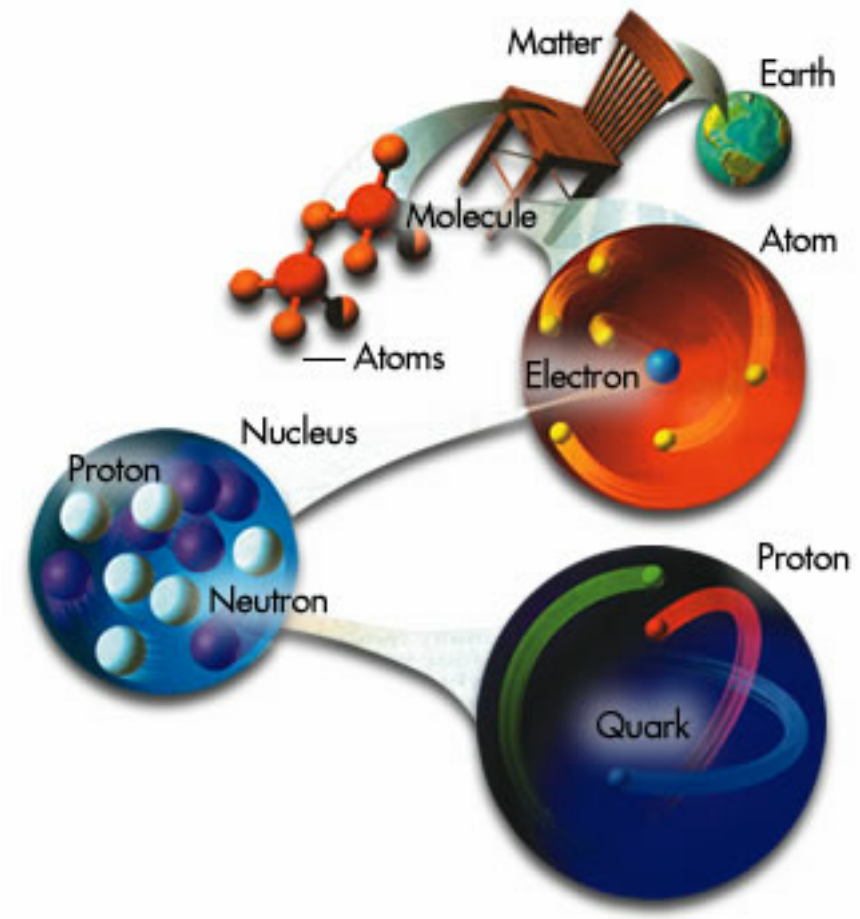
Terminology

Behaviors

Normalized Gauges

Gauges

Raw Data



What Data Was Captured?

- OWL ID
- Module #
- IU #
- Question #
- Session #
- Attempt #
- Score
- Answer Date
- Answer Time
- Seconds to Respond
- **UMass ID**
- **Question Type**
- **Due Date**
- **Due Time**

Items in **Boldface** were added later

Preprocessing

- Students with "Incomplete" grade removed
- Course split into engaged vs. disengaged
  - Engaged students attempted 85% of…
    - Homework assignments,
    - Lecture prep assignments,
    - PRS problems,
    - Course feedback surveys, and
    - Quizzes.
    - Attended all exams.
- Trial run indicated viability of study
- Each homework problem categorized

Problem Types

- Analysis
- Conceptual
- Multiple-Choice / Definition
- Traditional
- Problems were categorized by myself and Dr. Leonard

Gauges

- Calculated from raw data
- One specific measurement of student activity
  - Narrowly defined
  - Some seem like duplicates at first
- Usually a count or average
- Literal names:

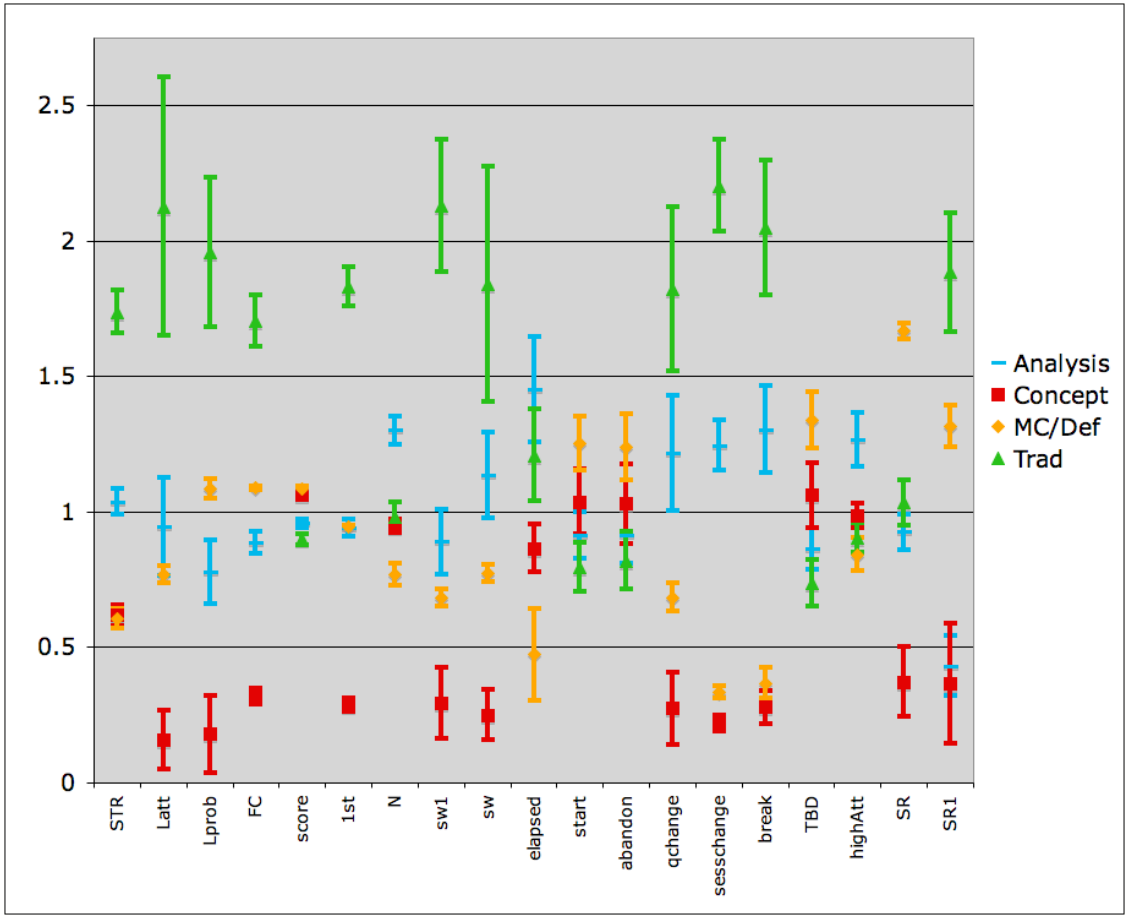
Seconds to Respond	Number of Attempts
Time Before Due	Start Time
Elapsed Time	Short Wrong
Late Problems	Credit per Attempt

What We Did With Gauges

- Verify validity of problem types
- Correlations with performance
  - Find predictors
  - Understand relations between problem types
  - Compare courses
- Correlations with each other
- Principal Component Analysis
- Combine to form Behaviors

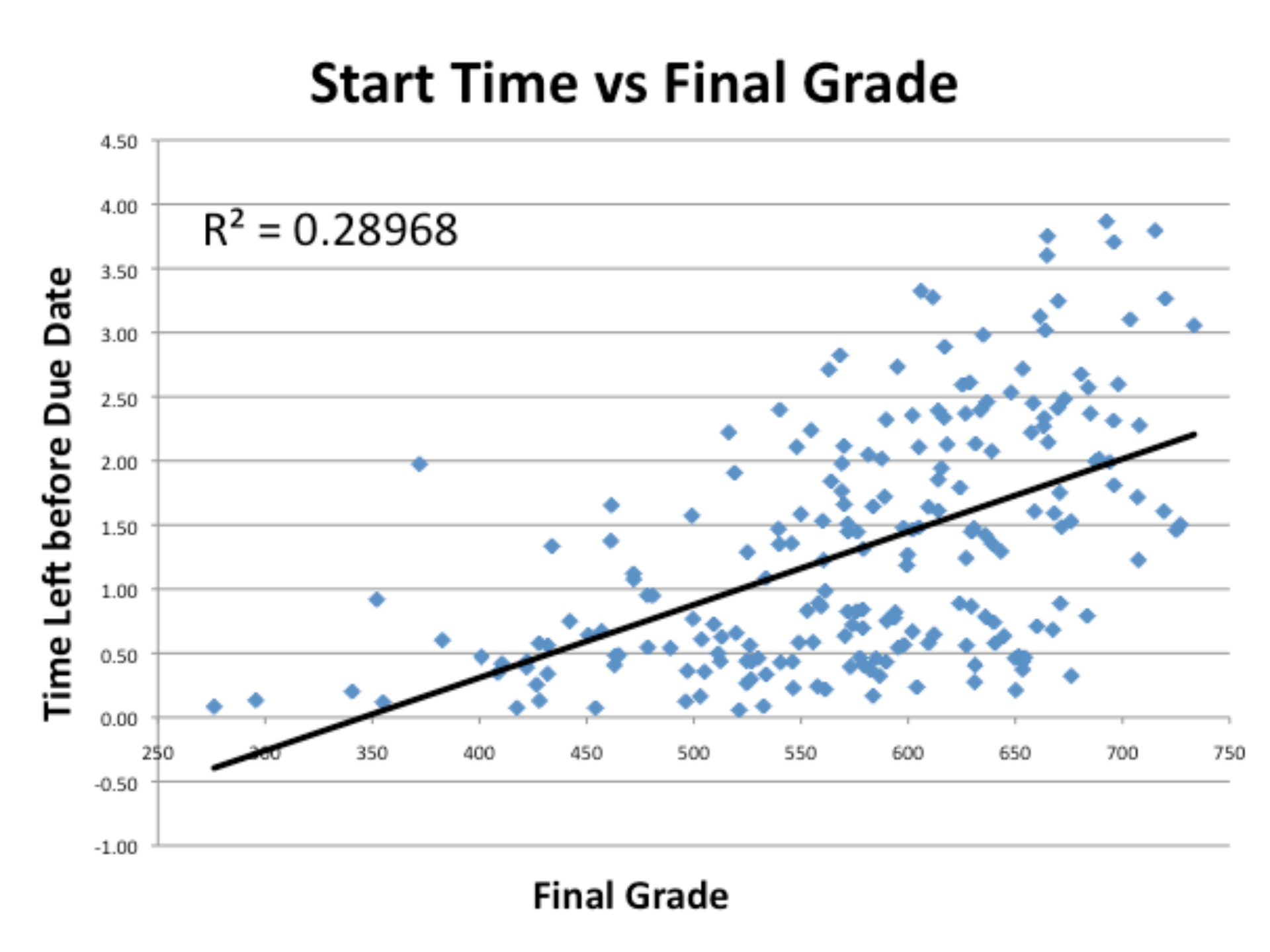
Problem Type Separation

- Gauges along bottom, arbitrary scale on left
- Error bars are twice standard error
- Excellent separation in many cases.

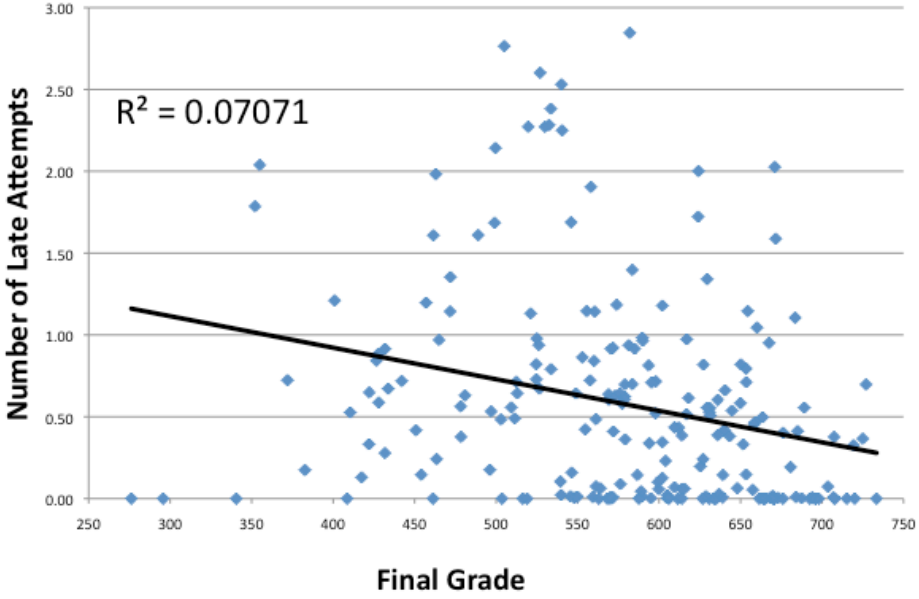


Gauges as Predictors

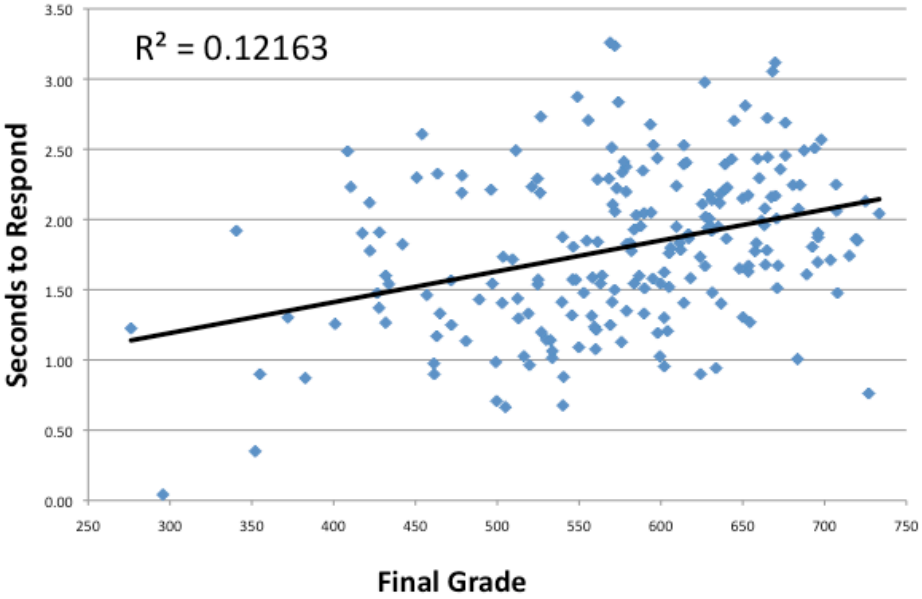
- Strong predictors:
  - Performance-based gauges (but not all)
  - Time-related gauges (but not all)
  - Attempt-related gauges (just about all)
- Courses often differed in gauge correlations.
- Behaviors make better predictors.


Start Time vs Final Grade
$R^2 = 0.28968$
Time Left before Due Date
Final Grade
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
250
350
400
450
500
550
600
650
700
750

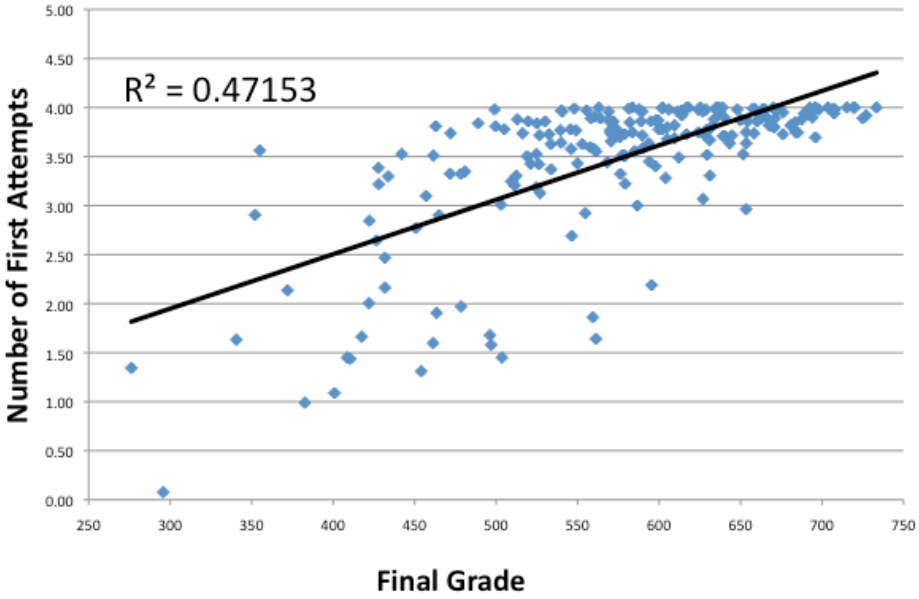
Late Attempts vs Final Grade



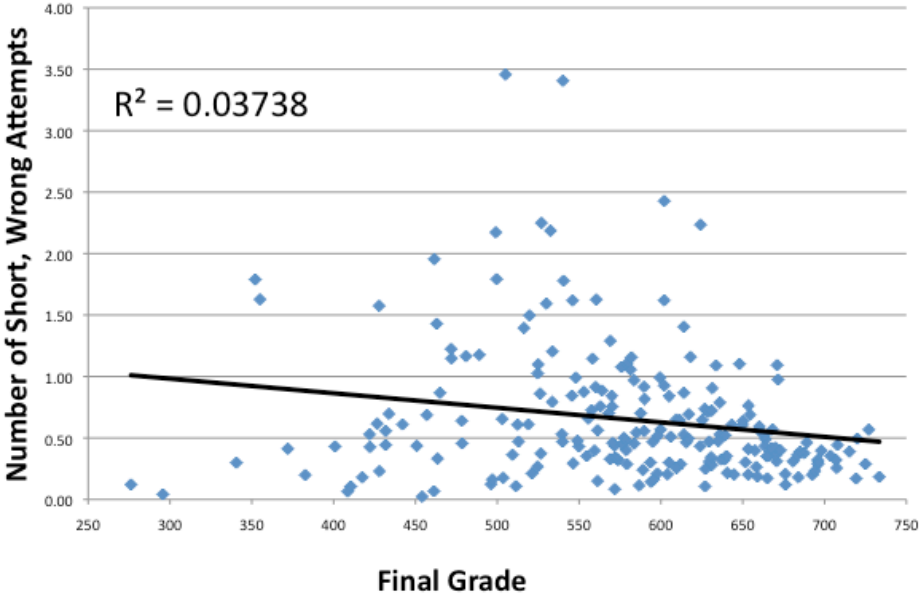
STR vs Final Grade



First Attempts vs Final Grade



SW vs Final Grade



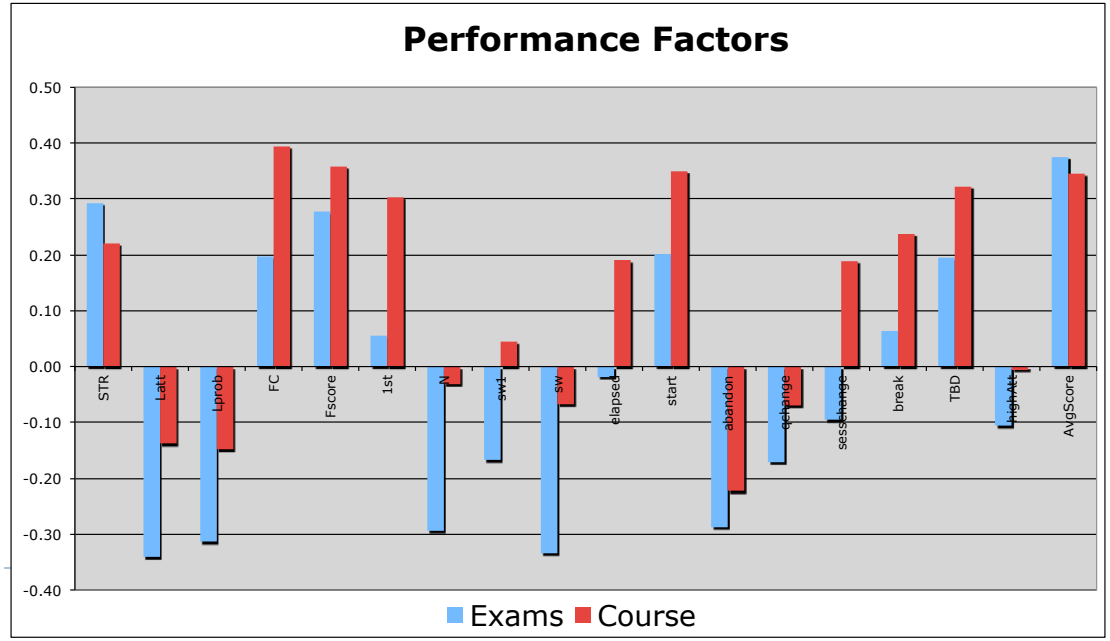
Gauge Cross-Correlation

- Groups created seem to be more functional than meaningful
- Not as powerful as PCA — no new factors, increased possibility of erroneous correlation

	STR	N	Latt	Lprob	qchar	sw1	sw	highA	sessi	break	1st	FC	Fscor	aban	AvgSc	elaps	start	TBD
STR	1	-0.4	-0.4	-0.4	-0.3	-0.5	-0.5	-0.3	-0.1	-0	0.1	0.14	0.21	-0.1	0.3	0.03	0.07	0.08
N	-0.4	1	0.77	0.68	0.7	0.71	0.68	0.7	0.51	0.39	0.37	0.35	0.14	-0.1	-0.4	0.1	-0	-0
Latt	-0.4	0.77	1	0.94	0.71	0.74	0.71	0.61	0.46	0.21	0.29	0.33	0.17	-0.1	-0.2	-0	-0.2	-0.2
Lprob	-0.4	0.68	0.94	1	0.73	0.64	0.67	0.54	0.49	0.25	0.31	0.31	0.14	-0.1	-0.2	0	-0.2	-0.2
qchange	-0.3	0.7	0.71	0.73	1	0.53	0.54	0.39	0.68	0.36	0.68	0.47	0.03	0.16	-0.3	0.16	-0	-0
sw1	-0.5	0.71	0.74	0.64	0.53	1	0.94	0.6	0.34	0.11	0.17	0.15	0.04	-0	-0.4	0	-0.1	-0.2
sw	-0.5	0.68	0.71	0.67	0.54	0.94	1	0.55	0.4	0.16	0.18	0.21	0.11	-0.1	-0.3	0.08	-0.1	-0.1
highAtt	-0.3	0.7	0.61	0.54	0.39	0.6	0.55	1	0.2	0.15	0.08	0.1	0.04	-0.1	-0.4	-0.1	-0.2	-0.2
sessions	-0.1	0.51	0.46	0.49	0.68	0.34	0.4	0.2	1	0.82	0.65	0.61	0.27	-0.1	0.05	0.28	0.27	0.22
break	-0	0.39	0.21	0.25	0.36	0.11	0.16	0.15	0.82	1	0.43	0.46	0.25	-0.2	0.05	0.31	0.33	0.27
1st	0.1	0.37	0.29	0.31	0.68	0.17	0.18	0.08	0.65	0.43	1	0.82	0.27	0.01	0.03	0.11	0.2	0.19
FC	0.14	0.35	0.33	0.31	0.47	0.15	0.21	0.1	0.61	0.46	0.82	1	0.74	-0.6	0.41	0.21	0.28	0.23
Fscore	0.21	0.14	0.17	0.14	0.03	0.04	0.11	0.04	0.27	0.25	0.27	0.74	1	-0.9	0.72	0.25	0.25	0.2
abandon	-0.1	-0.1	-0.1	-0.1	0.16	-0	-0.1	-0.1	-0.1	-0.2	0.01	-0.6	-0.9	1	-0.7	-0.2	-0.2	-0.1
AvgScore	0.3	-0.4	-0.2	-0.2	-0.3	-0.4	-0.3	-0.4	0.05	0.05	0.03	0.41	0.72	-0.7	1	0.1	0.22	0.21
elapsed	0.03	0.1	-0	0	0.16	0	0.08	-0.1	0.28	0.31	0.11	0.21	0.25	-0.2	0.1	1	0.69	0.67
start	0.07	-0	-0.2	-0.2	-0	-0.1	-0.1	-0.2	0.27	0.33	0.2	0.28	0.25	-0.2	0.22	0.69	1	0.98
TBD	0.08	-0	-0.2	-0.2	-0	-0.2	-0.1	-0.2	0.22	0.27	0.19	0.23	0.2	-0.1	0.21	0.67	0.98	1

PCA on Gauges

- Factors often differ between problem types
- Individual factors, esp. first ones, are often bad predictors. Good pred. come from many factors.
- Combos below: r=.4 for exams, r=.75 for course



Behaviors

- Linear combinations of Gauges
- Created through intuition and examination
- No “splitting hairs”
- Broader than Gauges
- Evocatively named

Sample Behaviors

- **Uncertainty**
  - STR +1
  - Attempts +1
  - sw +1
  - Qchange +1
  - Sessions +1
  - Breaks +1
- **Tenacity**
  - FC +2
  - Fscore –1
  - Problems +1
  - Abandon –1
- **Efficiency**
  - Attempts -1
  - sw1 -.5
  - sw -.5
  - AvgScore +1

More?

Get On With It.

Other Behaviors

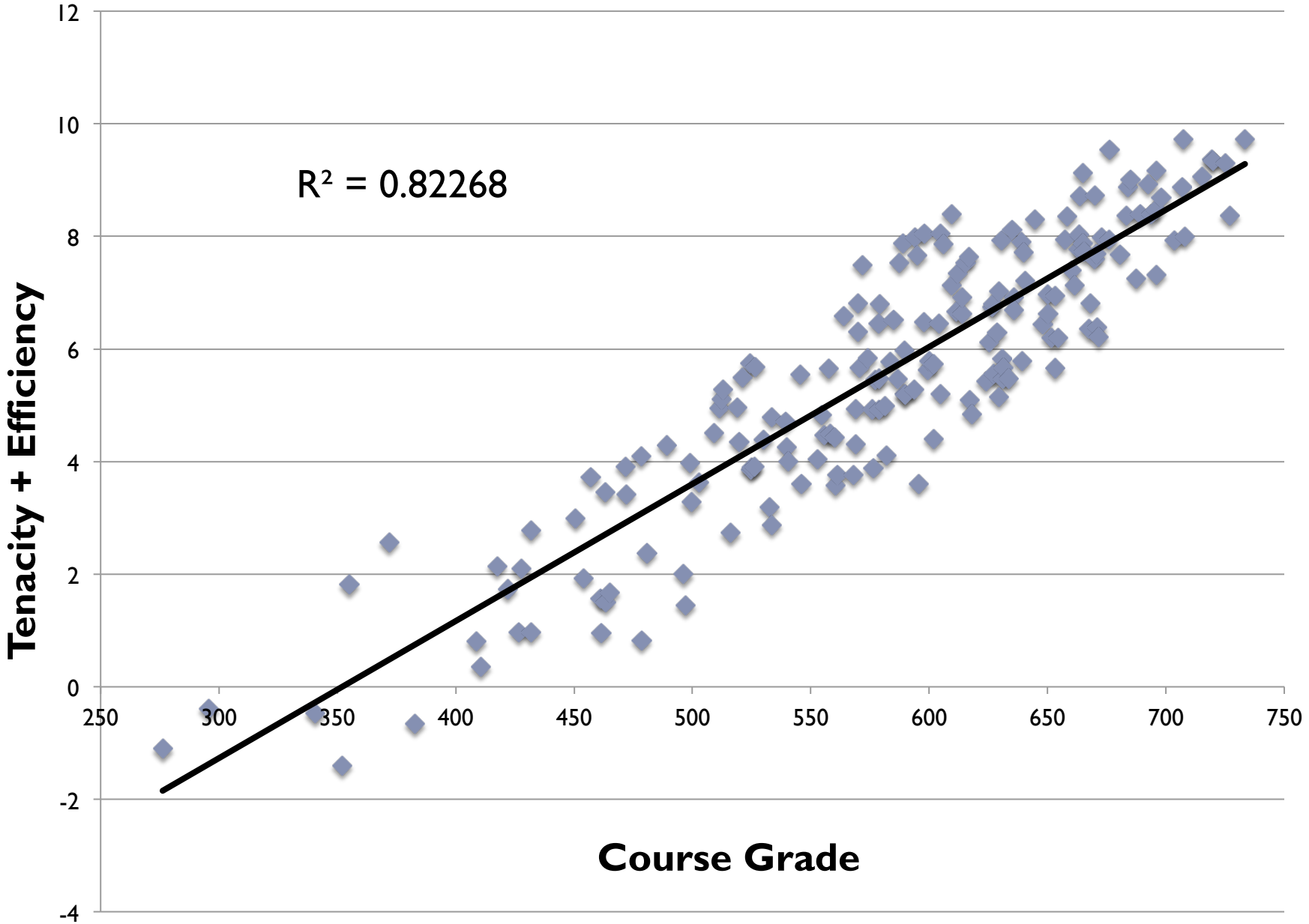
- Inactivity
  - Attempts –1
  - sw +1
  - sw1 +1
  - Start Time –1
  - TBD –1
  - Abandon +1
- Frustration
  - Attempts +1
  - sw +1
  - Qchange +1
  - Sessions +1
  - Abandon +1
- Slow & Steady
  - STR +1
  - Attempts +1
  - Elapsed +1
  - HighAtt +1
  - FC +1
- Grade-Conscious
  - Latt +1
  - LProb +1
  - Fscore +1
  - Problems +1

Tenacity & Efficiency

- Best predictor found
- Also easiest to explain! No fancy statistics, better for interpretation than PCA.

"r" values	T	E	T+E
P151 Exams	.32	.48	**.49**
P151 Course	.76	.32	**.80**
P181 Exams	.60	.48	**.71**
P181 Course	.81	.37	**.83**

Behavior and Grade, Physics 181 and 151



The Original Point: Analysis

- 181's Final Exam Question #3
  - A "How would you solve this problem?" question.
  - Frustration, inactivity, efficiency have no significant impact. Uncertainty is weak positive
  - Grade-conscious, tenacious, slow & steady are best
- "Fruitful Struggle" seems most effective
- Attempting many problems is as useful as getting high scores or starting early.
- Attempting analysis questions more worthwhile, despite higher number of traditional questions

Interesting 151 Survey Items

- Our time-related gauges do not match students' reporting of time spent

- Disengaged students often wanted to understand material; engaged prefer to improve existing knowledge

- Engaged students more likely to seek multiple resources when stuck
- Disengaged students more likely to give up or "keep trying"

Limitations of Methods & Data

- Behaviorist Bias
- Linear Modeling
- Noise, noise, noise
- Interpretation of higher-order constructs

Wrap-up

- Other approaches appearing in colleges
  - Degree Compass (predicts passing)
  - Course Signals (Nth Week Flag)
- Augmenting grading?
- Data from 8.011 / MITx?
  - Generalize to other disciplines?
  - Other gauges?
  - Longitudinal studies?

Acknowledgements

- William Gerace
- My Thesis Committee
- Ian Beatty
- Emma White

Future Work

- relate.mit.edu
- DontStopLearning.wordpress.com

Other Approaches and Methods (1)

- Behrouz Minaei-Bidgoli (2004, thesis, MSU)
  - Cluster analysis, genetic algorithms, pattern recognition, etc.
  - Optimized results account for ~90% of final grade
- Warnakulasooriya & Pritchard (2005, MIT)
  - Using gauges to classify problems by difficulty
- Kotas & Finck (2002, MSU)
  - Homework collaboration between students well-correlated with final grade
  - Surveys, log data, and institutional data

Other Approaches and Methods (2)

- Kortemeyer (2004, MSU)
  - "Effective Feedback to the Instructor from Online Homework"
- Cole and Todd (2003)
  - No significant difference between the performance of students using written or online homework, despite using "multimedia homework with immediate rich feedback."
  - Suspicions of bleed-through between experimental and control sections: students in pen-and-paper sections sometimes used the logins of the students in electronic homework sections in order to receive feedback

Research Younger Than Mine

- Lots more "e-homework works" papers, in various fields, especially finance/business.
- Butler, et. al. (2008, Mt. St. Mary's)
  - "… it was found that the students who received immediate feedback on quizzes had higher quiz and test averages than other students…"
- Kortemeyer (2009, MSU)
  - Gender differences in reported use
- Bennett, et. al. (2007, ???)
  - Data-Mining an Online Homework System