Engaging Faculty in Learning Analytics: from One Institution to Many

Tim McKay (@TimMcKayUM)
University of Michigan
Some background to set the stage

• Always drawn to projects with rich data and diverse goals (SDSS, ROTSE, DES):
  – Multidisciplinary teams who do what they must to get the data needed for big questions

• Why turn to education research now?
  – Rich, extensive data about learners and their contexts is newly available
  – A chance to make STEM education much more equitable and inclusive

• One important focus is on large, foundational courses
  – In practice: to design courses for equity, inclusion, and efficacy
  – In research: as laboratories for exploring teaching and learning within and across disciplines
Every year, thousands of our students need to be introduced to dozens of disciplines
University Hall was constructed in 1871. At a time when total enrollment at the University was 1200 students, this building featured an auditorium seating 3000. No small plans were made.
Early data at Michigan

- Universities have always used estimators of student achievement
  - At the course level
  - At the degree level
- From 1837 – 1912 only three course outcomes were used at Michigan:
  - Passed
  - Not passed
  - Conditioned
- Like diplomas, this is threshold testing rather than ranking

A mastery-based system...
Birth of the industrial university

• In 1900, enrollment had tripled, to 3482, and the industrial era had begun

• In 1950, enrollment expanded by an additional factor of ten, to 43,683

• Michigan became a model of a modern public research university

• Indoor graduation had become impossible... 1949 commencement on Ferry Field
The 20th Century began with an industrial revolution. Public higher education joined in: exploding in scale and adopted bureaucratic, industrial approaches, including standardized tests, credit hours, GPAs, majors, and minors.

The modern ‘academic record’ is a product of a higher education system deeply influenced by the industrial revolution. This ‘permanent record’ a precursor for today’s datafication of everything.
Academic transcripts: Designed to fit in an envelope

The transcript of the future is under no snail-mail constraints.

Let’s remember how thin this representation is, that it provides no context, does not record much of what happens in college...and fix it.
Grading as data reduction

• Grading is data reduction, inference from data...
  – Begin with the full body of a students work: homework, exams, papers, participation
  – Reduce this to a single estimator of student performance: a grade

• The whole of a semester recorded as one of 13 categories (A+ to E)

• **Ideally**, grades are rich, well informed, expert evaluations of authentic student work

• **In practice**, they are not always so...involving non-expert evaluation, mechanical assessment via multiple choice exams, systems which are easily circumvented

**Each student’s full four year experience is then further reduced to a single GPA**
Grades began at UM in 1912

<table>
<thead>
<tr>
<th>Class</th>
<th>Total</th>
<th>Left or Dropped</th>
<th>Incomplete from Examination</th>
<th>Total of Dropping</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-----------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>167</td>
<td>645</td>
<td>999</td>
</tr>
</tbody>
</table>

**Economics**

1. Taylor  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
2. Parry  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
3. Parry  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
4. Adams  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
5. Friday  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
6. Taylor  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
7. Parry  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
8. Cooley  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
9. Cooley  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7
10. Jones  2102  40  30  32  167  645  999  135  54  58.4  32.2  49.9  66.8  02.7

*Year 3 juniors, and 2 freshmen.*

Arthur C. Hall
Registrar

August 5, 1913.
Early 20th century grading

- Letters grades spread from the East Coast along with the Carnegie credit hour system
- Most schools moved to some form of grading system between 1900 and 1920
- In many cases, letter grades were intended to allow identification of strong students for groups like Phi Beta Kappa

- From the start, little attention was paid to what they mean or how to award them...
- Max Meyer in *The Grading of Students* (Science, 1908, 28, 243):
  “The experiment started by the faculty five years ago must be pronounced a complete failure. And both students and faculty have before now felt it to be a failure. There is no uniformity of grading, but the greatest divergence. It has come to be admitted openly that a student who is anxious to win honors must be careful to elect his work under certain teachers and avoid others as much as possible.”
Max Meyer in *The Grading of Students* (Science, 1908, 28, 243)

Data from the University of Missouri from 1902-1907, five years into their ‘experiment’ with letter grades...

Among these courses, the fraction of A grades varies by a factor of 50!
<table>
<thead>
<tr>
<th>Course name</th>
<th>25% Superior Students</th>
<th>50% Medium Students</th>
<th>25% Inferior Students</th>
<th>Total Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A B C</td>
<td>A B C D-F</td>
<td>A B C D-F</td>
<td></td>
</tr>
<tr>
<td>WOMENSTD 300</td>
<td>0.25 0 0</td>
<td>0.49 0.01 0 0</td>
<td>0 0.21 0.03 0.01</td>
<td>776</td>
</tr>
<tr>
<td>ASTRO 106</td>
<td>0.25 0 0</td>
<td>0.48 0.02 0 0</td>
<td>0 0.16 0.04 0.04</td>
<td>1275</td>
</tr>
<tr>
<td>GTBOOKS 191</td>
<td>0.25 0 0</td>
<td>0.46 0.04 0 0</td>
<td>0 0.24 0 0</td>
<td>2754</td>
</tr>
<tr>
<td>LING 111</td>
<td>0.25 0 0</td>
<td>0.39 0.11 0 0</td>
<td>0 0.17 0.05 0.03</td>
<td>2762</td>
</tr>
<tr>
<td>PHIL 180</td>
<td>0.25 0 0</td>
<td>0.28 0.22 0 0</td>
<td>0 0.11 0.09 0.05</td>
<td>3392</td>
</tr>
<tr>
<td>PSYCH 111</td>
<td>0.25 0 0</td>
<td>0.26 0.24 0 0</td>
<td>0 0.12 0.10 0.03</td>
<td>34114</td>
</tr>
<tr>
<td>CHEM 216</td>
<td>0.25 0 0</td>
<td>0.23 0.27 0 0</td>
<td>0 0.19 0.05 0.01</td>
<td>19653</td>
</tr>
<tr>
<td>ENGLISH 125</td>
<td>0.25 0 0</td>
<td>0.22 0.28 0 0</td>
<td>0 0.19 0.05 0.01</td>
<td>39934</td>
</tr>
<tr>
<td>PSYCH 240</td>
<td>0.25 0 0</td>
<td>0.21 0.29 0 0</td>
<td>0 0.09 0.12 0.04</td>
<td>7582</td>
</tr>
<tr>
<td>EARTH 105</td>
<td>0.25 0 0</td>
<td>0.19 0.31 0 0</td>
<td>0 0.06 0.14 0.06</td>
<td>387</td>
</tr>
<tr>
<td>POLSCI 101</td>
<td>0.25 0 0</td>
<td>0.16 0.34 0 0</td>
<td>0 0.18 0.06 0.02</td>
<td>4111</td>
</tr>
<tr>
<td>EEB 390</td>
<td>0.25 0 0</td>
<td>0.16 0.34 0 0</td>
<td>0 0.08 0.13 0.04</td>
<td>1124</td>
</tr>
<tr>
<td>BIOLOGY 173</td>
<td>0.25 0 0</td>
<td>0.15 0.35 0 0</td>
<td>0 0.17 0.07 0.01</td>
<td>7593</td>
</tr>
<tr>
<td>CHEM 230</td>
<td>0.25 0 0</td>
<td>0.14 0.36 0 0</td>
<td>0 0.09 0.14 0.03</td>
<td>6950</td>
</tr>
<tr>
<td>ENGR 100</td>
<td>0.25 0 0</td>
<td>0.14 0.36 0 0</td>
<td>0 0.15 0.09 0.01</td>
<td>11852</td>
</tr>
<tr>
<td>STATS 250</td>
<td>0.25 0 0</td>
<td>0.13 0.37 0 0</td>
<td>0 0.02 0.16 0.07</td>
<td>7667</td>
</tr>
<tr>
<td>PHYSICS 135</td>
<td>0.25 0 0</td>
<td>0.11 0.39 0 0</td>
<td>0 0 0.21 0.04</td>
<td>2347</td>
</tr>
<tr>
<td>EECS 280</td>
<td>0.25 0 0</td>
<td>0.05 0.42 0.04 0</td>
<td>0 0 0.17 0.08</td>
<td>5881</td>
</tr>
<tr>
<td>PHYSICS 235</td>
<td>0.25 0 0</td>
<td>0.04 0.38 0.07 0</td>
<td>0 0 0.19 0.06</td>
<td>1731</td>
</tr>
<tr>
<td>ECON 102</td>
<td>0.25 0 0</td>
<td>0.04 0.38 0.08 0</td>
<td>0 0 0.18 0.07</td>
<td>13526</td>
</tr>
<tr>
<td>ECON 101</td>
<td>0.25 0 0</td>
<td>0.04 0.36 0.10 0</td>
<td>0 0 0.17 0.08</td>
<td>28537</td>
</tr>
<tr>
<td>CHEM 215</td>
<td>0.25 0 0</td>
<td>0.02 0.43 0.05 0</td>
<td>0 0 0.18 0.07</td>
<td>20472</td>
</tr>
<tr>
<td>MATH 215</td>
<td>0.25 0 0</td>
<td>0.01 0.40 0.09 0</td>
<td>0 0 0.17 0.08</td>
<td>19996</td>
</tr>
<tr>
<td>MCDB 310</td>
<td>0.25 0 0</td>
<td>0 0.44 0.06 0</td>
<td>0 0 0.13 0.12</td>
<td>2019</td>
</tr>
<tr>
<td>ECON 401</td>
<td>0.24 0.01 0</td>
<td>0 0.35 0.15 0</td>
<td>0 0 0.13 0.12</td>
<td>9717</td>
</tr>
<tr>
<td>CHEM 210</td>
<td>0.23 0.02 0</td>
<td>0 0.40 0.10 0</td>
<td>0 0 0.15 0.10</td>
<td>20561</td>
</tr>
<tr>
<td>PHYSICS 240</td>
<td>0.23 0.02 0</td>
<td>0 0.34 0.16 0</td>
<td>0 0 0.20 0.05</td>
<td>13644</td>
</tr>
<tr>
<td>BIOLOGY 172</td>
<td>0.22 0.03 0</td>
<td>0 0.36 0.14 0</td>
<td>0 0 0.14 0.11</td>
<td>6317</td>
</tr>
</tbody>
</table>

**BIOLOGY 171** | 0.21 0.04 0 | 0 0.42 0.08 0 | 0 0 0.17 0.08 | 7386
**MATH 116** | 0.21 0.04 0 | 0 0.38 0.12 0 | 0 0 0.16 0.09 | 25344
**PHYSICS 140** | 0.21 0.04 0 | 0 0.32 0.18 0 | 0 0 0.18 0.07 | 20284
**CHEM 130** | 0.21 0.04 0 | 0 0.42 0.08 0 | 0 0 0.18 0.07 | 20334
**MCDB 427** | 0.18 0.07 0 | 0 0.35 0.15 0 | 0 0 0.13 0.12 | 923
**MATH 115** | 0.15 0.10 0 | 0 0.36 0.14 0 | 0 0 0.15 0.10 | 24543

Data from the University of Michigan, from 1996-2012.

Courses selected for study by the 2013 Learning Analytics Fellows Program

Among these courses, the fraction of A grades varies by a factor of 4.9
Clarence S. Yoakum: VP for Educational Investigations
PRE-COLLEGE PROGRESS AND COLLEGE GRADES

Office of Educational Investigations

January, 1931

The following data are selected from a mass of material in the files which relate to the progress of students, insofar as that can be gauged by grades. The relation in later semesters is not high found is probably low. New for it deals with the large at the greatest point of st record is also included.

CORRELATIONS SHOWING HOW WELL UNIVERSITY GRADES PREDICT OTHER UNIVERSITY GRADES

<table>
<thead>
<tr>
<th>Subjects</th>
<th>1926</th>
<th>1927</th>
<th>1928</th>
<th>1929</th>
</tr>
</thead>
<tbody>
<tr>
<td>English - 1st Sem. with 2nd Sem.</td>
<td></td>
<td></td>
<td></td>
<td>.580</td>
</tr>
<tr>
<td>Math - 1st Sem. with 2nd Sem.</td>
<td></td>
<td></td>
<td></td>
<td>.452</td>
</tr>
<tr>
<td>Mod.For.Lang. - 1st Sem. with 2nd Sem.</td>
<td></td>
<td></td>
<td></td>
<td>.660</td>
</tr>
<tr>
<td>Total - 1st Sem. with 2nd Sem. Engineers</td>
<td></td>
<td></td>
<td>.700</td>
<td>.696</td>
</tr>
<tr>
<td>Total - 1st Sem. with 2nd Sem. L. S. &amp; A.</td>
<td>.680</td>
<td>.706</td>
<td>.663</td>
<td></td>
</tr>
<tr>
<td>Total - 1st Sem. with 2nd Sem. Freshmen</td>
<td></td>
<td>.647</td>
<td>.669</td>
<td>.678</td>
</tr>
<tr>
<td>Total - 1st Sem. with 2nd Sem. Sophomores</td>
<td></td>
<td></td>
<td>.682</td>
<td></td>
</tr>
<tr>
<td>Total - 1st Sem. Fresh. with 1st Sem. Soph.</td>
<td></td>
<td></td>
<td>.557</td>
<td></td>
</tr>
<tr>
<td>Total - Freshmen Year with Sophomore Year</td>
<td></td>
<td></td>
<td></td>
<td>.675</td>
</tr>
</tbody>
</table>
The total result indicates most strongly that there is a large element of chance in the selection of individuals for college entrance. This result might be expected, since grades for different years in college correlate for the most part only in the fifties.

The study covered some ten thousand individuals over a decade. It is one of the most comprehensive investigations on record in the number of indi-
A pause while data languished

- Fricke’s report marked the end of what began with Yoakum and others in the first decades of the 20th century.

Beginning in the 1970’s, higher ed worked hard to more deeply personalize, with richer advising, freshman seminars, learning communities, and more.

Doing this at scale is difficult and costly. So what has been done reaches most students very thinly...
Teaching & Learning data became increasingly extensive and accessible

- UM Student Record Data in the Warehouse: Begun in 1996, our institutional repository
- Academic information about over 100,000 students: from admission to graduation
- A small record of expensively processed, high S/N data
- Usage and student work from online tools: LMS, Google tools, Blogs, Piazza, Coursera, etc.
- Activity Records in support systems like the library, science learning center, housing, student groups, UROP, etc.
- Huge, noisy data sets, but potentially very rich
Cautions about Datafication

• Data systems regularly represent individuals with reductive labels: gender, first gen status, GPA, major, residency, race
  – Each collapses a more complex identity to a single *measurable type* (Cheney-Lippold 17)
  – We might use single quotes to distinguish reductive measurable types (e.g. ‘gender’), from the complex social identities (*gender*) they mean to represent

• This is an inescapable challenge for data science: representations are not reality.
What insights are lost with such reductive representations?
Re-emergence of LA for UM...

- The USE lab: Stephanie Teasley, Steve Lonn, and many colleagues
- Initially focused on LMS interaction data:
  - learner-content
  - learner-learner
  - learner-instructor
- Early explorations of digitally mediated education – clickstreams, interaction data, activity
- Several efforts emerging from physics
- Gus Evrard, LSA ITS, and ART 1.0: 2001/2002
  - Access to information in the UM data warehouse
- McKay, Evrard, and Gerdes: the better-than-expected project 2008+
  - Efforts to understand how preparation affects performance in large introductory courses

In November of 2010, Stephanie Teasley and I met to discuss exploring a larger learning analytics community at Michigan and launched the SLAM series
The SLAM seminars

- We aim to bring together people interested in understanding our academic mission through analysis of data
- We are a Rackham Distinguished Faculty and Graduate Student Seminar

- And an M-Community group: learning-analytics@umich.edu

Bi-weekly seminars

- September 14: Timothy McKay, UM Physics
  49,000 Physics Students: Who Does Better Than Expected?
- September 28: Rebecca Marx & Mark Banaszak Holl UM Chemistry
  Concurrent Enrollment in Lecture and Laboratory Enhances Student Performance and Retention
- October 12: John Campbell, Associate Vice President for IT, Purdue
  "Signals": The Past, Present, and Future Application of Analytics
- October 26: Anne Gere, UM Education and English
  Quantitative Measures of Writing Ability
- November 9: David Pritchard, MIT Physics
  Patterns, Correlates, and Reduction of Homework Copying

Volunteer to talk/send suggestions to McKay
Provost’s Learning Analytics Task Force launched May 2012

Committee of 12 faculty

The LATF charge

1. Optimize the UM information environment for learning analytics
2. Fund a series of the best proposed LA projects at UM
3. Review the metrics used to assess teaching and learning at UM

Accomplishments

• During the task force
  – SLAM: learning from data
  – ELA Grants
  – LA Fellows Program
  – Teaching evaluation project
• New things created
  – Digital Innovation Greenhouse: acting on data
  – Practical LA Coursera MOOC
  – Learning Analytics Data Architecture
  – UMILA
Provisioning of data for research: Michigan’s LARC system

A ‘regular release’ model for clean research data. Similar to those in open science projects like the SDSS or GAIA space mission.
Two results from dozens that came from LA Fellows...both with important implications for equity.
Michigan 1:2:1 Introductory Chemistry Curriculum Model:

- Structure and Reactivity I (organic chemistry) with Laboratory
  - below 70th percentile chemistry and below 30th percentile math

- Physical Chemistry Principles and Applications
  - above 70th percentile chemistry and 30th percentile math
  - biology, non-science pre-meds or pre-health

Traditional 2:2 Introductory Chemistry Curriculum Model:

- 2 Semesters General Chemistry
- 2 Semesters Organic Chemistry

Chemistry Placement Analysis
Placement by Ethnicity

- ▲ - Reported AP above 2
- ○ - Placed Orgo, no AP

Chem Placement

2 or More  | Asian      | African American | Hispanic | Not Indicated | White
0.00  | 0.10  | 0.20  | 0.30  | 0.40  | 0.50  | 0.60
#2: Learning from the past

• What happened in the past allows us to define *better-and-worse than expected* performance. Doing this is not prediction, it is learning from experience in an effort to change the future...

• Doing this, we learn new things. Here are two:
  – All intro STEM lecture courses impose a *grade penalty* on students
  – Significant *gender performance disparities* are apparent in these classes too: men and women with the same preparation have different outcomes
1st lesson: substantial grade penalties are imposed by all intro STEM lecture courses

Average grade anomaly:

\[ \text{AGA} = \langle \text{Grade} - \text{GPAO} \rangle \]

\[ \langle \text{Grade} - \text{GPAO} \rangle_{\text{all}} = -0.45 \]
Gendered performance differences

2nd lesson: substantial gendered performance differences exist

\[
\begin{align*}
<\text{Grade} - \text{GPAO}>_{\text{male}} &= -0.32 \\
<\text{Grade} - \text{GPAO}>_{\text{female}} &= -0.59 \\
\text{GPD} &= \text{AGA}_{\text{female}} - \text{AGA}_{\text{male}} = -0.27
\end{align*}
\]
This raised a lot of questions

- Is this just a problem in physics, or does it affect other disciplines as well?
- Which ones? What is it about the disciplines or classes that makes this happen?
- Is this just a problem at Michigan, or are they seen on many campuses? Which ones? Why?

For higher education to advance, it’s not enough to analyze our data in single contexts. We must share the results of parallel analyses across & between institutions.
Higher mean grades

Lower mean grades

Females favored

Males favored

STEM Lectures

STEM Lab Courses

Patterns of AGA and GPD at Michigan

The punchline: It’s not just physics

Details: Koester et al., ArXiv 1608:07565
Sloan Learning and Research Analytics (LARA) Project: 2015

- LARA was a trial for larger collaborations
- Parallel analysis on many Big Ten campuses
  1. Regularize data on each campus
  2. Share analysis code which runs on regularized data sets
  3. Compare results across campuses
- Many involved: 52 people from 10 Big Ten institutions
- **One example analysis:**
  - Are the patterns of grade penalty and gender difference seen in large courses at Michigan widespread?
  - Five institutions (incl. Indiana), 172 courses, 677,949 students

Many thanks to key organizers Becky Matz (MSU) and Ben Koester (Umich)!
Comparing PHYS-I-E

Male/Female difference at Michigan: 0.23

Male/Female diff: 0.22

Modest but very significant gender differences in Physics

Male/Female diff: 0.15

Male/Female diff: 0.24
Comparing ECON-I

Male/Female difference at Michigan: 0.32

Male/Female diff: 0.31

Male/Female diff: 0.30

Big gender differences in Economics on all the campuses we examined

Male/Female diff: 0.36
Comparing Chem-III

Male/Female difference at Michigan: 0.18

Male/Female diff: 0.22

Smaller, but still very significant differences in Organic Chemistry

Male/Female diff: 0.14
Comparing Chem-I-Lab

Male/Female difference at Michigan: 0.01

Male/Female diff: -0.01

Virtually no difference in any of the lab classes

Male/Female diff: -0.08

Male/Female diff: 0.09
These patterns of low grades and gendered performance difference appear on large university campuses across the country.

The punchline: It’s not just Michigan

Next generation LA projects

• Broader data use within campuses
  – Research access for the community should be as easy as possible
  – Tools which use data to help students should be explored
  – Much richer descriptions of student experiences are needed by all

• Sharing of information among universities
  – We must be able to understand and explain the impacts of structures we impose
  – We have to provide rich, robust evidence about how college affects long-term life outcomes
Drawing together local STEM reform efforts from ten large, public research universities currently enrolling a more than 350,000 students.
College & Beyond II

• In the late ’90s, Bill Bowen led the College & Beyond study
  – 32 selective colleges and universities, data about more than 100,000 from 1951, 1976, and 1989
  – Survey data on later education, occupations, life satisfaction, civic involvement, and retrospective views of college

• Now launching College and Beyond II led by Paul Courant & ICPSR at UM
  – Pilot of 6-8 large public universities, data from cohorts 2000, 2005, 2010
  – Much richer measures of college experience: breadth, depth, effort, diversity of experience...
  – Follow-up with both administrative and survey data on life outcomes