Next Steps for Big Data in Education

Chris Dede
Harvard University
Chris_Dede@harvard.edu
www.gse.harvard.edu/faculty/christopher-dede
My Current Role in Data-Intensive Research in Education

• Organized a two workshop sequence on data-intensive research for NSF and the field: insights from relatively mature data-intensive research initiatives in the sciences and engineering were applied to nascent data-intensive research efforts in education

• Confront “big data” issues in my design-based research in ecosystems science education
Definitions

• Big Data is characterized by the ways in which it allows researchers to do things not possible before (i.e., Big data enables the discovery of new information, facts, relationships, indicators, and pointers that could not have been realized previously).

• Data-intensive research involves data resources that are beyond the storage requirements, computational intensiveness, or complexity that is currently typical of the research field.

• Data science is the large-scale capture of data and the transformation of those data into insights and recommendations in support of decisions.
BIG DATA

→ VOLUME
→ VELOCITY
→ VARIETY
→ VERACITY
From Hindsight to Foresight

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics

Value vs. Difficulty:
- Hindsight
- Insight
- Optimization
- Foresight
From Description to Prescription

• Determine students’ probabilities of failure (*predictions*)
• Determine which students respond to which interventions (*uplift modeling*)
• Determine which interventions are most effective (*explanatory modeling*)
• Allocate resources accordingly (*cost benefit analysis*)
PAR Framework

ACT

Student watch lists for targeted interventions

Reusable predictive models

Measurable Intervention results

Common Definitions of terms

Common Definitions of interventions

Multi-Institutional collaboration

PREDICT

RETAIN

Scalable, cross-institutional improvements that support each individual student’s success
PAR Data Inputs

Student Demographics & Descriptive
- Gender
- Race
- Prior Credits
- Perm Res Zip Code
- HS Information
- Transfer GPA
- Student Type

Student Course Information
- Course Location
- Subject
- Course Number
- Section
- Start/End Dates
- Initial/Final Grade
- Delivery Mode
- Instructor Status
- Course Credit

Student Financial Information
- FAFSA on File – Date
- Pell Received/Awarded – Date

Student Academic Progress
- Current Major/CIP
- Earned Credential/CIP

Course Catalog
- Subject
- Course Number
- Subject Long
- Course Title
- Course Description
- Credit Range

Lookup Tables
- Credential Types Offered
- Course Enrollment Periods
- Student Types
- Instructor Status
- Delivery Modes
- Grade Codes
- Institution Characteristics

Possible Additional
- Placement Tests
- NSC Information
- SES Information
- Satisfaction Surveys
- College Readiness Surveys
- Intervention Measures

** Future
### Descriptive and Predictive Insight

**PAR Benchmarks**

**Descriptive Analytics**

**Cross Institutional**

Student/degree/major level insight into:

1. What **did** the retention look like for students entering in the same cohort?
2. How does your institution compare to peer institutions/institutions in other sectors?
3. How **did** performance vary by student attributes?

### PAR Models

**Predictive Analytics**

**Institutional Specific** insight into:

1. What students are being retained over time?
2. Which students are currently at risk for completing and why?
3. Which factors are directly correlated to student success?
4. What is the predicted course completion rate for a particular program?
Outcome Measures

• Credits Earned
• Credits Earned- No Dev Ed
• Credit Ratio
• Credit Ratio- No Dev Ed
• Retention
• Credentials Earned (Any, Bachelors, Associates, Certificate)
Predicting retention aimed at taking action - finding the most important factors
Key Research Questions

• Can we detect problems that students are having as they are happening, through automated analysis?

• Can we provide real-time feedback to students and educators in response to the problem detection?

• Is the feedback effective in helping students attain more sophisticated behaviors? Does it make sense to the students and educators? Is it actionable in that they are able to do something useful with it?
Key Next Steps

• Mobilize Communities around Opportunities based on New Forms of Evidence
• Develop New Forms of Educational Assessment
• Develop New Types of Analytic Methods
• Build Human Capacity to Do and to Understand Data Science
• Develop Advances in Privacy, Security, and Ethics
• Infuse Evidence-based Decision-Making throughout Organizations and Systems
3 E’s of Immersive Learning

- **Engagement**
  Students are motivated to do well, see the relevance of their learning, and increase in self-efficacy

- **Evocation**
  Immersive interfaces can evoke a wide spectrum of authentic performances with embedded support

- **Evidence**
  Log files, chat logs, shared notebooks, and similar artifacts provide a rich evidentiary trail
EcoMUVE – Multi-User Virtual Environment
Collaborative construction of concept maps
Augmenting Real World Ecosystems

http://ecomobile.gse.harvard.edu

(Conner Flynn)
GoPro Cameras Capture EcoMOBILE Experience
Now we need to write two things we see that could affect dissolved oxygen.

Plants.

Plants. And...you guys would rather say...rain, or the dead matter?

Mm, dead matter, maybe. 'Cause—

(Why the) dead matter is bacteria?

—Yeah. The bacteria. And we don't know, you know, how long...this has been...

You got plants already? Plants, 'cause they release dissolved oxygen into the water.

This could...

Provide food for bacteria, increasing their population and increasing their need for dissolved oxygen.

The bacteria and—

And um—

And causing an increase in population.

Yeah, increasing their population and their need for dissolved oxygen.

[Student talking to other student] Quinn.

Um, provide food for bacteria, increasing population?

Mhm. [Partner 1 continues typing in Evernote]

Plants could release dissolved oxygen into the water and dead matter could provide food for bacteria, increasing the bacteria population and their need for dissolved oxygen.
EcoMUVE

- MUVEs promote self-efficacy in science
- Simulate experiences otherwise impossible in school settings.
- Explore time and scale
- Opportunities to take on roles, work in teams
- Shared immersive experience that contextualizes learning and supports inquiry

(Ketelhut et al. 2010, Metcalf et al. 2011)

EcoMOBILE

- Greater fidelity and sensory richness, physical interactions with organisms and environments.
- Self-directed collection of real-world data and artifacts.
- Facilitated use of cameras, recording devices, probes, GPS, mapping, graphing, augmented reality.
What Can We Inculcate and Assess?

- Inquiry skills?
- Collaboration?
- Leadership?
- Self-efficacy?
- Metacognition?
Key Research Questions

- Can we detect problems that students are having as they are happening, through automated analysis?
- Can we provide real-time feedback to students and educators in response to the problem detection?
- Is the feedback effective in helping students attain more sophisticated behaviors? Does it make sense to the students and educators? Is it actionable in that they are able to do something useful with it?
From Hindsight to Foresight
Imagining the Possible

Edith S. Gummer
Ewing Marion Kauffman Foundation

NSF Advancing Data Intensive Research in Education
June 1-2 2015
### District and School Information

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>District Accreditation</td>
<td>Raw Data: Current accreditation status for districts.</td>
</tr>
<tr>
<td>District Calendar Days and Hours</td>
<td>Raw Data: Comparison of district’s calendar information.</td>
</tr>
<tr>
<td>School District by Representative District</td>
<td>List of school districts by Missouri Representative District.</td>
</tr>
<tr>
<td>School District by Senatorial District</td>
<td>List of school districts by Missouri senatorial district.</td>
</tr>
<tr>
<td>School District by Congressional District</td>
<td>List of school districts by United States Congressional District.</td>
</tr>
<tr>
<td>School District with their United States Congressional District</td>
<td>List of School Districts with their United States Congressional District.</td>
</tr>
<tr>
<td>School District Map - County Boundaries</td>
<td>Map of Missouri school districts with county boundaries.</td>
</tr>
<tr>
<td>School District Map - House District Boundaries</td>
<td>Map of Missouri school districts with House district boundaries.</td>
</tr>
<tr>
<td>School District Map - Senate District Boundaries</td>
<td>Map of Missouri school districts with Senate district boundaries.</td>
</tr>
<tr>
<td>School District Map - Congressional District Boundaries</td>
<td>Map of Missouri school districts with congressional district boundaries.</td>
</tr>
<tr>
<td>District Enrollment by Grade</td>
<td>Raw Data: District enrollment by grade level for 1991-2014 (data as of 10/31/2014)</td>
</tr>
<tr>
<td>District Enrollment 2007-2012</td>
<td>Raw Data: Multi-year enrollment by district.</td>
</tr>
<tr>
<td>District Enrollment 2014</td>
<td>Raw Data: District enrollment by district – ranked by percent of total state enrollment.</td>
</tr>
</tbody>
</table>

### District and Building Graduation and Dropout Indicators

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>District Annual Dropout Rate</td>
<td>Raw Data: District dropout information. Includes total dropouts and race breakdowns.</td>
</tr>
<tr>
<td>District Graduate Analysis</td>
<td>Raw Data: District graduate analysis. Includes placement status on previous year’s graduates.</td>
</tr>
<tr>
<td>District Graduation Rate</td>
<td>Raw Data: District graduation rates. Includes total graduates and race breakdowns.</td>
</tr>
<tr>
<td>Four-Year District Graduation Rate</td>
<td>PEP: Four-Year State and District Graduation Rates</td>
</tr>
<tr>
<td>Building Annual Dropout Rate</td>
<td>Raw Data: Building dropout information. Includes total dropouts and race breakdowns.</td>
</tr>
<tr>
<td>Building Graduate Analysis</td>
<td>Raw Data: Building graduate analysis. Includes placement status on previous year’s graduates.</td>
</tr>
<tr>
<td>Building Graduation Rate</td>
<td>Raw Data: Building graduation rates. Includes total graduates and race breakdowns.</td>
</tr>
<tr>
<td>Graduates 16 Month Postsecondary Follow-up (SPSFO-16)</td>
<td>Report: Number of graduates who have enrolled in Postsecondary education within 16 months of graduation. This report is being published as an</td>
</tr>
</tbody>
</table>

### District and Building Student Indicators

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>District Attendance Rate</td>
<td>Raw Data: District attendance rate. Includes K-12 and 9-12 rates.</td>
</tr>
<tr>
<td>District Demographic Data</td>
<td>Raw Data: District demographic data. Includes enrollment information by race. Data as of 10/31/2014</td>
</tr>
<tr>
<td>District Discipline Incident</td>
<td>Raw Data: District discipline incidents. Includes enrollment, number of incidents, and rates by offense and removal types.</td>
</tr>
<tr>
<td>District Proportional Attendance Rate</td>
<td>Raw Data: District proportional attendance rate (students attending 90% of the time). Years 2009-2013. Includes K-9 and 9-12 rates.</td>
</tr>
<tr>
<td>Building Attendance Rate</td>
<td>Raw Data: Building attendance rates.</td>
</tr>
<tr>
<td>Building Demographic Data</td>
<td>Raw Data: Building demographic data. Includes enrollment information. Data as of 10/31/2014</td>
</tr>
<tr>
<td>Building Discipline Incident</td>
<td>Raw Data: Building discipline incidents. Includes enrollment, number of incidents, and rates by offense and removal types.</td>
</tr>
<tr>
<td>Building Proportional Attendance Rate</td>
<td>Raw Data: Building proportional attendance rate (students attending at least 90% of the time). Years: 2009-2013. Includes K-9 and 9-12 rates.</td>
</tr>
</tbody>
</table>

### District and Building Education Staff Indicators

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>District Faculty Information</td>
<td>Raw Data: District faculty information. Includes salary, degree, and years of experience. Data as of 10/31/2014</td>
</tr>
<tr>
<td>District Staff Ratio</td>
<td>Raw Data: District staff ratios. Includes student to teachers and student to administrator ratios. Data as of 10/31/2014</td>
</tr>
<tr>
<td>Building Faculty Information</td>
<td>Raw Data: Building faculty information. Includes salary, degree, and years of experience. Data as of 10/31/2014</td>
</tr>
<tr>
<td>Building Staff Ratio</td>
<td>Raw Data: Building staff ratios. Includes student to teachers and student to administrator ratios. Data as of 10/31/2014</td>
</tr>
<tr>
<td>Building Certification</td>
<td>Raw Data: Building teacher certification information. Includes number of teachers with a valid certificate and highly qualified teachers.</td>
</tr>
<tr>
<td>District Certification</td>
<td>Raw Data: Building teacher certification information. Includes number of teachers with a valid certificate and highly qualified teachers.</td>
</tr>
</tbody>
</table>
The Data Behind EdWise

• 14 million records
• 600+ metrics
• Aggregate only, no PII
• Nine themes
• 100% of data is K-12

Assessment, 28.2%
Finance, 27.7%
Teacher & Staff, 13.1%
Completion & Matriculation, 10.3%
Demographics, 7.4%
Community Indicator, 4.5%
Discipline, 3.8%
Transportation, 3.8%
Accreditation, 1.4%
Tool 1: Mapping
Tool 2: Find a District/School

<table>
<thead>
<tr>
<th>Metric Measurement Level</th>
<th>Beginning of School Year</th>
<th>State</th>
<th>Grade/Assessment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>(All)</td>
<td>(All)</td>
<td>Null</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric Category [A]</th>
<th>Metric Value [A]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Enrollment</td>
<td>10000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric Category [B]</th>
<th>Metric Value [B]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students Per Teacher</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric Category [C]</th>
<th>Metric Value [C]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of Students with Free or Reduced Lunch, January</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric Category [D]</th>
<th>Metric Value [D]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of Non-White Students within the Fall Enrollment</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Key Research Questions

• Can we detect problems that students are having as they are happening, through automated analysis?

• Can we provide real-time feedback to students and educators in response to the problem detection?

• Is the feedback effective in helping students attain more sophisticated behaviors? Does it make sense to the students and educators? Is it actionable in that they are able to do something useful with it?
From Hindsight to Foresight

- **Value**
  - Descriptive Analytics
  - Diagnostic Analytics
  - Predictive Analytics
  - Prescriptive Analytics

- **Difficulty**
  - Hindsight
  - Insight
  - Optimization

- **Gartner**
Key Next Steps

• Mobilize Communities around Opportunities based on New Forms of Evidence
• Develop New Forms of Educational Assessment
• Develop New Types of Analytic Methods
• Build Human Capacity to Do and to Understand Data Science
• Develop Advances in Privacy, Security, and Ethics
• Infuse Evidence-based Decision-Making throughout Organizations and Systems
Questions to Ponder

• To what types of behavioral data could we now apply these methods?
  – *Micro-level* data (e.g., each student’s second-by-second behaviors as they learn)
  – *Neso-level* data (e.g., teachers’ patterns in instruction; students’ patterns in retention)
  – *Macro-level* data (e.g., aggregated student outcomes for accountability purposes) Gummer’s work with EdWise

• What are the barriers to collecting, storing, sharing and analyzing these data?

• How can we build human and organizational capacity to use evidence-based findings effectively?