Education Leadership Data Analytics: 
*Integrating Education Leadership, Data Science, and Evidence-based Improvement Cycles in Schools*

Alex J. Bowers, Ph.D.

Associate Professor of Education Leadership  
Teachers College, Columbia University  
New York, New York, USA  
http://www.tc.columbia.edu/faculty/ab3764/  
http://www.tc.columbia.edu/elda/  
Twitter: @Alex_J_Bowers  
ORCiD ID: 0000-0002-5140-6428
Overview: Education Leadership Data Analytics (ELDA)

- AI in education, data analytics, and education data science. Hype versus reality.
  - What is Education Leadership Data Analytics (ELDA) and why should anyone care?

- Data science, and education leadership training

- What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?
  - Interaction and small group discussion

- Pattern visualization of student course interactions:

- Calculating predictor accuracy for early warning systems

Understanding How Educators Use Instructional Data Warehouses (IDWs) (or not) NSF collaboration:

- Building community and capacity in data use: Nassau BOCES and Teachers College collaborative

- Do teachers and administrators have different perceptions of data use in their schools and IDW use?
Growing interest in AI in Education

Artificial Intelligence, Authentic Impact: How Educational AI is Making the Grade

Educators find AI can revolutionize the K-12 experience, for both teachers and students.

AI Applications In Education

Ron Schmelzer, Contributor

AI becoming an increasingly part of education industry

As artificial intelligence becomes an increasing part of our daily lives, it’s no wonder that educational institutions are racing to catch up with the need to develop more talented, development-ready curricula, but the education being transformed is institutions from elementary to high school and well as adult and post-secondary. Educators are transforming by integrating AI to help students learn better and meet educational objectives.

14 Questions Educators Should Ask About AI-Based Products

Michele Meinar, Associate Editor

What Every Educator Needs to Know About Artificial Intelligence

Experts think artificial intelligence could help students on all aspects of their learning, from self-driving cars, to science, and even, learning K-12 education. Only.

Teachers, the Robots Are Coming. But That’s Not a Bad Thing

By Kewin Backhalter

January 7, 2020

Bring up the idea of even the possibility of artificially intelligent robots replacing some of what teachers do, and you are likely to spark a torrent of anger among many educators. Intelligent machines could never match human interactions, they argue. Such moves would be a giant step toward a digital dystopia in education.

That kind of reaction to the role of AI robots in education clearly played out in our recent Big Ideas survey of K-12 teachers, which featured questions about robotics. The vast majority of teachers, 84 percent, disagreed with the suggestion that student learning would likely improve if more K-12 teachers had AI-powered robots working with them as classroom assistants. More than 90 percent did not think that student learning would improve in classrooms where chronically low-performing human teachers were replaced by artificially intelligent robots.
2019 was the highest on record $1.7 Billion

Source:
EdSurge Jan 15, 2020
EdSurge Dec 19, 2017
But what is AI in Education?
But what is AI in Education?

Robots looking past chalk boards?
The Difference between Statistics and Machine Learning

Data → Statistics → Outcome

Model → (Known beforehand)
The Difference between Statistics and Machine Learning

**Statistics**
- Data
- Model
- (Known beforehand)
- Outcome
  - (Identified at the end)

**Machine Learning**
- Data
- Outcome
  - (Known beforehand)
- Model
  - (Identified at the end)
The Difference between Statistics and Machine Learning

Statistics

Data → Model → Outcome (Identified at the end)

Model (Known beforehand)

Machine Learning

Data → Outcome (Known beforehand)

Model (Identified at the end)

Predict

Outcome

Lots more Data

This is the Power of AI.
Rapid improvements over the last few years in machine learning and deep learning neural networks driven initially through access to ImageNet datasets.

![Examples from CIFAR-10 dataset](image)

Figure 1.1: Examples from CIFAR-10 dataset

The Change in Easy to Implement Neural Networks for Image Recognition Over Just 1.5 years

Enjoying learning #DeepLearning in @AndrewYNg's @coursera MOOC. Training a #NeuralNet for Cat/Not Cat in Python deeplearning.ai

76% probability cat (2017)
Data stored in local drive
Analysis done on local machine (bare metal)
The Change in Easy to Implement Neural Networks for Image Recognition Over Just 1.5 years

Enjoying learning #DeepLearning in @AndrewYNg's @coursera #MOOC. Training a #NeuralNet for Cat/Not Cat in Python deeplearning.ai

76% probability cat (2017)
Analysis on my computer

98% probability frog (2019)
Data stored in the cloud
Analysis done with Docker, Keras, Tensorflow

Really fun to create a "frog" detection #DeepLearning network w/ #Tensorflow #Keras Python code from Douglas Blank's @Cometml #ODSCEast warmup. Great #ModelVis tools! So cool how fast domain is developing! #MachineLearning #DataScience #Analytics nbviewer.jupyter.org/github/Calysto...
The problem

- Image recognition is something humans do well already.

- What about AI for things we’re not very good at? Such as outcome prediction in education?

This talk:
- Bringing together learning analytics, machine learning, and education leadership for evidence-based improvement
- K-12 education dropout risk prediction
- Different types of dropouts? Typology analytics
- Data dashboards, educator data use, and school improvement
Early Warning Prediction/Dashboards in K-12 Education

Panorama

Illuminate

Infinite Campus
One school district
65,000 students
670 teachers
73 schools

Relationship of Instructional Data Warehouse (IDW) “Instructional Clicks” to Student Achievement
Weak relationship with elementary reading
No relationship with elementary or junior high math, or junior high reading over 3 years

Longitudinal Effects of Teacher Use of a Computer Data System on Student Achievement

Jeffrey C. Wayman
Wayman Services, LLC
Shana Shaw
Texas Center for Educator Excellence
Vincent Cho
Boston College

Does data use make a difference in student achievement? Despite the field's optimism on this matter, relatively few studies have attempted to quantify the effects of data use. These studies have often used the presence of a data use intervention (e.g., a data system or data coaching) as a proxy for use, as opposed to tracking teachers’ direct interactions with data, via data system click logs, for example. Accordingly, the present study sought to address this methodological gap by exploring the 2-year effects of data use through a multilevel cross-classified model of teachers’ system interactions and student achievement. A significant relationship was found between system use and elementary reading, but no significant relationships were found for elementary math, junior high math, or junior high reading. The implications of this study on how to conceptualize and measure use, as well as how to support practitioners, are discussed.

https://doi.org/10.1177/2332858416685534
Early Warning System Dashboard Randomized Controlled Experiments

Main Finding = No effect

Journal of Research on Educational Effectiveness
Volume 12, 2019 - Issue 3

An Efficacy Study of a Ninth-Grade Early Warning Indicator Intervention

Martha Abiele Mac Iver, Marc L. Stein, Marcia H. Davis, Robert W. Balter & Joanna Hornig Fox

Pages 363-390 | Received 18 May 2018, Accepted 12 April 2019, Published online 31 Jul 2019

Abstract

Building on previous research showing how well ninth-grade student behaviors predict on-time high school graduation, this experimental study investigates the impact of a ninth-grade intervention on student attendance and course passing. The study, conducted in 41 geographically and demographically diverse high schools within a single state, evaluates the effects of placing a half-time staff member in high schools to implement the Early Warning Intervention (EWI) Team model designed to monitor ninth-grade early warning indicators and provide timely interventions. Analyses based on the pre-specified student outcomes of attendance rate and percentage of ninth-grade course credits earned indicated no statistically significant impact of the intervention. On secondary outcome variables, results indicated that students in treatment schools were significantly less likely than control school students to be chronically absent. The difference between treatment and control school students on dichotomous measures of course failure were not statistically significant. The widespread dissemination of research and best practices related to early warning systems and ninth-grade interventions likely accounted for low levels of contrast between treatment and control school practices and outcomes.


https://doi.org/10.1080/19345747.2019.1615156
Machine Learning for High School Predictors of Graduation
Models not generalizable = Must re-analyze in three different districts

“In short, the best indicators of failure to graduate on time for one district and grade level may not be the best for other districts and grade levels”

Issues in Education Predictive Analytics & Early Warning Systems

• Early Warning Systems (EWS) and Indicators (EWI) randomized controlled trials have to date shown little effect on student persistence.
  – Admittedly, the evidence is sparse.

• Early evidence suggests that machine learning prediction algorithms may not generalize across contexts well.

• To date, interventions that target student K-12 persistence don’t show much impact (Freeman & Simonsen, 2015)

• So what does a data analyst in education do?

Central issues for Education Leadership Data Analytics (ELDA) (and today’s talk):

• Build capacity of people who can talk to both the machines and to people: Education Data Science
• Build stronger collaborations between dashboard and analytics providers and educators, figuring out what data will actually support instructional improvement.
• Increase the accuracy of prediction and dashboard systems.
• Compare what educators say they want with data versus what they actually click on.
• Build collaborations between education systems, researchers, and analytics vendors.
We have reason to fear that the multitude of books which grows every day in a prodigious fashion will make the following centuries fall into a state as barbarous as that of the centuries that followed the fall of the Roman Empire.

- Adrien Baillet, 1685


Big Data:

- **Volume**  – Scale of Data
- **Variety**  – Different forms of Data
- **Velocity** – Analysis of Data
- **Veracity** – Uncertainty of Data

You’re dealing with Big Data when you’re working with data that doesn’t fit into your computer unit. Note that makes it an evolving definition: Big Data has been around for a long time. . . . Today, Big Data means working with data that doesn’t fit in one computer. p.24
“Big” Analysis in Education Leadership

Mplus
Statistical Analysis With Latent Variables
HLM
Hierarchical Linear & Nonlinear Modeling
R
RapidMiner
Python
Tableau
Stata Data Analysis and Statistical Software
Schools have used algorithms for a long time
What is Data Science?
What is Data Science?

Drew Conway ODSC East 2018 – “Data scientists are people who can write on glass backwards”
A data scientist is someone who knows how to extract meaning from and interpret data, which requires both tools and methods from statistics and machine learning, as well as being human… Once she gets the data into shape, a crucial part is exploratory data analysis, which combines visualization and data sense… She’ll communicate with team members, engineers, and leadership in clear language and with data visualizations so that even if her colleagues are not immersed in the data themselves, they will understand the implications. p.16
The 26 Steps of Doing Data Science

1. Clean data.
2. Clean data.
3. Clean data.
4. Clean data.
5. Clean data.
6. Clean data.
7. Clean data.
8. Clean data.
9. Clean data.
10. Clean data.
11. Clean data.
12. Clean data.
13. Do some math.
14. Try to get everyone to understand my findings.
15. Try to get everyone to understand my findings.
16. Try to get everyone to understand my findings.
17. Try to get everyone to understand my findings.
18. Try to get everyone to understand my findings.
19. Try to get everyone to understand my findings.
20. Try to get everyone to understand my findings.
21. Try to get everyone to understand my findings.
22. Try to get everyone to understand my findings.
23. Try to get everyone to understand my findings.
24. Try to get everyone to understand my findings.
25. Try to get everyone to understand my findings.
26. Repeat.

Data scientist surpasses statistician on Google Trends

The relative interest in data scientist surpassed statistician this month. It was also higher in April and September of this year, so it's not new, but it does seem like it's ready to be a consistent thing, at least least for a little while. That said, it doesn't seem like statistician is losing interest to data scientist, as the former has been fairly consistent for the past few years, so take that how you want.

http://flowingdata.com/2013/12/18/data-scientist-surpasses-statistician-on-google-trends/
Leading with Evidence in Schools: Data and Research Literacy

March 2 - 29, 2020

Upcoming course offerings:
Spring Session - March 2 - 29, 2020
Summer Session - July 6 - August 2, 2020
Fall Session - November 2 - 29, 2020

Where: Online asynchronous course

Registration Fee: $595

Group/Team Discount (5 or more):
25% off the registration fee, please contact cps@tc.columbia.edu to register

School administrators and educators employed by the New York City Department of Education are eligible for individual discounts of 25%. Please contact cps@tc.columbia.edu for more information.

Units Awarded:
Participants receive 20 Clock Hours and 20 CTLEs (applicable only to NYS residents).

http://tc.columbia.edu/cps/evidence
Quantitative Research Methods Training in Education Leadership and Administration Preparation Programs as Disciplined Inquiry for Building School Improvement Capacity

Alex J. Bowers

Abstract
The quantitative research methods course is a staple of graduate programs in education leadership and administration. Historically, these courses serve to train aspiring district and school leaders in fundamental statistical research topics. This article argues for programs to focus as well in these courses on helping aspiring leaders develop skills as practitioner-scholars, including deepening their practice around data analytics, providing opportunities to read and evaluate peer-reviewed research, analyzing data using current methods, and applying findings to facilitate building evidence-based improvement cycles in their schools. Additional data leadership training should be offered for the practicing administrator, educational quantitative analyst, research specialist, and district data scientist.
Training School System Leaders for Four Different Data Analytic Roles

Primary Quantitative Program
Focus on evidence-based improvement cycles
- Quantitative research methods
- Journal clubs
- University-district partnerships

Evaluation & Analysis Focus
- Advanced inferential statistics
- Psychometrics & testing
- Data management & ethics
- Survey & program evaluation
- Cost-benefit analysis
- Accounting & budgeting

Educational Quantitative Analyst

Research Specialist

Data Scientist

Organization-Level Data Analytics
- Education data mining
- Learning analytics
- Programming
- Technology & instruction
- Design-based research

http://doi.org/10.1177/1942775116659462
Make visible data that have heretofore gone unseen, unnoticed, and therefore unactionable. (p.ix) Bienkowski, Feng, Means, (2012)
Abstract:
Education Leadership Data Analytics (ELDA) is an emerging domain that is centered at the intersection of education leadership, the use of evidence-based improvement cycles in schools to promote instructional improvement, and education data science. ELDA practitioners work collaboratively with school and district leaders and teachers to analyze, pattern, and visualize previously unknown patterns and information from the vast sets of data collected by schooling organizations, and then integrate findings in easy to understand language and digital tools into collaborative and community-building evidence-based improvement cycles with stakeholders. In June of 2018, over 100 participants gathered for the Education Leadership Data Analytics Summit at Teachers College, Columbia University in New York City, including researchers, practitioners, policymakers, and funders. This report provides a summary of the central issues, themes, and recommendations for the future of the field that emerged from the discussions at the ELDA Summit event. These issues include building capacity in the field through incentivizing researcher practitioner partnerships, and providing conference and networking opportunities, professional development, certification, and ultimately degree programs to train ELDA researchers and practitioners. Additionally, a central focus of the ELDA field is equity, data security and privacy, in concert with open and FAIR data standards to develop and share de-identified data and tools across contexts. We conclude the report with a blueprint of possible skills and competencies needed for ELDA practitioner training and professional development and provide recommendations for next steps to help grow the field.

**ELDA Definition:**
Education Leadership Data Analytics (ELDA) practitioners work collaboratively with schooling system leaders and teachers to analyze, pattern, and visualize previously unknown patterns and information from the vast sets of data collected by schooling organizations, and then integrate findings in easy to understand language and digital tools into collaborative and community-building evidence-based improvement cycles with stakeholders.

**Identified Needs/Challenges:**

- Collaborative partnerships with schooling organizations
- Capacity-building and training infrastructure
- Focus on equity and algorithmic fairness
- Data privacy and security
- Open and accessible data and tools using F.A.I.R. data standards
Data Analytics & Education Data Science


Tommaso Agasisti
Politecnico di Milano, Italy
Figure 1. A data analytics model in education

Source: Authors’ elaborations, originally inspired by Siemens (2013).

Potential recommendations or decisions with algorithms: *The code must be open access! Taxpayers paid for it.*
Overview: Education Leadership Data Analytics (ELDA)

• AI in education, data analytics, and education data science. Hype versus reality.
  • What is Education Leadership Data Analytics (ELDA) and why should anyone care?

• Data science, and education leadership training

• What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?
  • Interaction and small group discussion

• Pattern visualization of student course interactions:

• Calculating predictor accuracy for early warning systems

Understanding How Educators Use Instructional Data Warehouses (IDWs) (or not) NSF collaboration:

• Building community and capacity in data use: Nassau BOCES and Teachers College collaborative
  • Building Community and Capacity for Data-Intensive Evidence-Based Decision Making in Schools and Districts. National Science Foundation, NSF DGE-1560720: [Link]

• Do teachers and administrators have different perceptions of data use in their schools and IDW use?
  • [Link]
What do teachers, principals and superintendents say they want in a longitudinal data system?

Longitudinal Data Use: Ideas for District, Building, and Classroom Leaders

INTRODUCTION
When educators effectively access and use data to drive their instructional planning, they are more responsive to student needs, more collegial in their interactions, and more reflective in their instructional practices (Hamilton et al., 2009; Mandich & Gummer, 2013; Wayman, 2005; Wayman & Conoley, 2008; Wayman & Springfield, 2008). Educators who frequently use student data can examine, expose, and understand patterns in student performance that allow them to better shape the instructional path for each student and meet goals across the district, thereby allowing for more responsive educational practices (Bowen, 2008; Mandich, 2012; Wayman & Springlight, 2008). This study examines how these types of school leaders—superintendents, principals, and teachers—perceive the purposes of longitudinal data use. The purpose of this study was to inform the development of a statewide longitudinal data system (SLDS) in a southeastern state in the United States. In this qualitative descriptive study, researchers prompted participants to respond with their wishes for a longitudinal-data tool specifically designed to meet their particular needs. This prompt was designed not only to inform the implementation of an SLDS system but to assist the researchers with understanding the practice of data use for teachers, principals, and superintendents. As Coburn (2012) noted, research about data use is a relatively new and emerging area and the study of data use has limited investigations related to the practices of school leaders. The data collected from these three participants groups allowed researchers to examine the practice of using longitudinal data sets for instructional improvement. This study sought to answer the following research question using the narrative data from school leaders’ responses:

What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal, and teacher leaders?

LITERATURE
The use of longitudinal school data is clearly documented in the professional research and trade literature. Perspectives from practitioners and researchers are various. The literature in this piece frames our study using the lens of leaders at the school-district superintendent level, the school-building principal level, and the school-classroom teacher level. This literature review presents published findings about the important aspects of a longitudinal school data system for these three, distinct types of school leaders.

The Superintendent and Longitudinal Data Use
Compared to teachers and principals, superintendents have the potential to use data in a broader capacity to better understand how not only to shape practices within schools but also to shape the vision that they and other stakeholders share for what the schools in an entire district should become. Mandich (2012) noted that superintendents have a responsibility to generate a districtwide commitment to use data and clear expectations about the purpose of such data use, and Bowen (2008) stressed that successful schools have “an overall system that integrates form and function while channeling the efforts of the employees toward a common goal” (p. 3). Decman et al. (2010) interviewed 16 superintendents from the Southern Gulf Coast region of Texas about their views on connecting the Intermediate School Leaders/Licensure Consortium Standard One to practice in their districts. In their interviews, the researchers determined 11 themes, including data-driven decision making, but they found that fewer than 15% of superintendents in the study gathered data from all stakeholders when building shared vision in their districts even though 50% said that data should be used to drive programmatic and instructional decisions. Ultimately, Decman et al. (2010) concluded that although the demographics of the sampled districts allows for generalizing to other regions, more data is needed, specifically data on how superintendents spend their time on various tasks. The researchers claimed that data could “clarify the difference between what superintendents say they value and how they actually spend their time” and “be useful for correlating superintendent behavior with self-reports” (p. 24). Although this study may not have provided a rich picture of superintendents’ use of data to create a shared vision, they did report that using data to inform practice, which raises the question about whether the school administrators and superintendents are merely using the specific data that is available or whether they are
What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?
What components of a longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal, and teacher leaders?
Discuss with One or Two People Around You…

Imagine the average teacher in your schools and channel their thoughts…

1. If you had a single webpage only you could visit to get quick information about students or schools, what information would the page contain?

2. In an ideal world, what information should the longitudinal data tool provide to teachers to help them make better decisions?

Repeat for the average principal and then repeat again for the Superintendent.
Discuss your responses with the people around you:

1. What were the similarities and differences between your answers for teachers?
2. What about for principals?
3. What about for the superintendent?
4. If you were to think about a Venn Diagram of your answers.
   - What are the overlaps?
   - What is unique for the superintendent, principals or teachers?
What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?
Superintendents \((n=55)\)

**Individual Student/Teacher**
- Teacher data: performance, attendance, past experience
- Who rides the bus
- Record of parent conferences
- Test scores for students as well as whether or not they have met growth

**Comparative**
- All district’s comparative test data, district demographics
- Access test data, budget information, comparing per pupil expenditures, demographic data for the local area as compared to the region and state
- Student growth from year to year, demographic group growth
Principals \((n=45)\)

**Teacher Data**
- Teachers’ past performance and years of experience
- Background of teachers and records showing student test scores broken down by teacher
- Student and teacher achievement and growth data, teacher evaluation information, student attendance and dropout data, staff attendance, accountability information

**Student Data**
- Achievement information on teachers based on student performance
- Background of teachers being hired; teacher performance data
- Are we able to keep high performing teachers in our district?
An overwhelming number of teacher data needs were related to a teacher’s own students, classroom, or personal performance. Nearly 96%, or 178 data points, could be categorized in this way. Resoundingly, teachers wanted longitudinal data that was individually connected to their own work. Teachers professed to wishing for longitudinal data that could be accessed by each teacher for each student on demand from an electronic portal, such as a classroom computer terminal or a remotely accessible portal.
What components of a longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal, and teacher leaders?


Figure 1. Summary of eight themes and corresponding frequency percentage in the “more frequent” level of category
Overview: Education Leadership Data Analytics (ELDA)

AI in education, data analytics, and education data science. Hype versus reality.

What is Education Leadership Data Analytics (ELDA) and why should anyone care?

Data science, and education leadership training


What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?

Interaction and small group discussion

Pattern visualization of student course interactions:


Calculating predictor accuracy for early warning systems


Understanding How Educators Use Instructional Data Warehouses (IDWs) (or not) NSF collaboration:

Building community and capacity in data use: Nassau BOCES and Teachers College collaborative


Do teachers and administrators have different perceptions of data use in their schools and IDW use?


Variables
Persons
Clinicians
Occasions x Variables
“diagnosis”
Clinicians
Occasions x Variables
“diagnosis”

Educators
Persons x Occasions
“growth”
Hierarchical Cluster Analysis Heatmaps and Pattern Analysis: An Approach for Visualizing Learning Management System Interaction Data

Ji Eun Lee  
Utah State University  
jieun.lee@aggiemail.usu.edu

Mimi Recker  
Utah State University  
mimi.recker@usu.edu

Alex J. Bowers  
Columbia University  
bowers@tc.columbia.edu

Min Yuan  
University of Utah  
min.yuan@eccles.utah.edu

ABSTRACT
This paper presents a form of visual data analytics to help examine and understand how patterns of student activity—automatically recorded as they interact with course materials while using a Learning Management System (LMS)—are related to their learning outcomes. In particular, we apply a data mining and pattern visualization methodology in which usage patterns are clustered using hierarchical cluster analysis (HCA) then visualized using heatmaps to produce what is called a clustergram. We illustrate the application of this methodology by building two clustergrams in order to explore university students’ LMS activity patterns using both semester and weekly summary data. The resulting clustergrams reveal differences in LMS usage between high-achieving and low-achieving/dropout students.

average users’ behaviors, and thus make it difficult to recognize the diverse patterns displayed by different groups of users [6]. Thus, in the present study, we take a more person-centered approach to visually investigate what sub-groups of students may share common patterns, and how these relate to their learning outcomes.

2.2 Hierarchical Cluster Analysis Heatmaps
HCA is a multivariate statistical method for classifying related units in an analysis across high dimensionality data. More recently, HCA has been combined with heatmap visualizations, called a clustergram [1]. The clustergrams represent each participant’s row of data across each of the columns of variables as a color block, using stronger intensities of one color to...
Canvas Learning Management System LMS Data
Mid-sized University
Undergrad freshman mathematics course, taught completely online as a required course
n=139 students
Clickstream LMS logfile data
Features are # of pageviews

**Method:**
Hierarchical Cluster Analysis (HCA) Heatmaps

Clusters together students with similar longitudinal data patterns and visualizes similarities and differences Bowers (2007, 2010)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Definition</th>
<th>M</th>
<th>Std.</th>
<th>Min/Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>announcements_v</td>
<td># of visits to announcement page</td>
<td>1.83</td>
<td>3.85</td>
<td>0 - 35</td>
</tr>
<tr>
<td>assignment_v</td>
<td># of times assignment viewed</td>
<td>11.86</td>
<td>7.11</td>
<td>0 - 31</td>
</tr>
<tr>
<td>assignment_p</td>
<td># of times assignment submitted</td>
<td>3.35</td>
<td>1.84</td>
<td>0 - 5</td>
</tr>
<tr>
<td>grades_v</td>
<td># of visits to grades page</td>
<td>11.34</td>
<td>11.02</td>
<td>0 - 62</td>
</tr>
<tr>
<td>attachment_v</td>
<td># of times attachment viewed</td>
<td>24.81</td>
<td>27.59</td>
<td>0 - 86</td>
</tr>
<tr>
<td>syllabus_v</td>
<td># of visits to syllabus page</td>
<td>7.25</td>
<td>6.34</td>
<td>0 - 33</td>
</tr>
<tr>
<td>discussion_v</td>
<td># of times discussion viewed</td>
<td>8.14</td>
<td>7.32</td>
<td>0 - 62</td>
</tr>
<tr>
<td>discussion_p</td>
<td># of times participated in discussion</td>
<td>1.23</td>
<td>0.89</td>
<td>0 - 6</td>
</tr>
<tr>
<td>wiki_v</td>
<td># of times wiki viewed/edited</td>
<td>70.87</td>
<td>58.49</td>
<td>0 - 295</td>
</tr>
</tbody>
</table>
Cluster Analysis Heatmaps of Canvas Interaction Data

Pattern by Content

Annotation column is final grade: from red (high) to dark (low)

Student are rows

Course features are columns
Cluster Analysis Heatmaps of Canvas Interaction Data

Pattern by Content

Pattern by Number of Interactions per Week

Student are rows

Course features are columns
Cluster Analysis Heatmaps of Canvas Interaction Data

Pattern by Content

Pattern by Number of Interactions per Week
Accurately Predicting K-12 Schooling Outcomes

- Dropout risk
  - Accuracy, precision, sensitivity & specificity
- Early Warning Indicators (EWI) / Early Warning Systems (EWS)
- Data driven decision making (3DM)
- Targeted and data-informed resource allocation to improve schooling outcomes


Issues of Accuracy in Predicting High School Dropout

- Dropping out of high school in the U.S. is associated with a multitude of negative outcomes.

- However, other than a select number of demographic and background variables, we know little about the accuracy of current dropout predictors that are school-related.

- Current predictions accurately predict only about 50-60% of students who actually dropout.

- According to Gleason & Dynarski (2002), accurate prediction of who will dropout is a resource and efficiency issue:
  - A large percentage of students are mis-identified as at-risk.
  - A large percentage of students who are at-risk are never identified.

- Current reporting of dropout “flags” across the literature is haphazard.
  - Almost none report accuracy
  - Many report specificity or sensitivity, but rarely both

- A dropout flag may be highly precise, in that almost all of the students with the flag dropout, but may not be accurate since the flag may identify only a small proportion of the dropouts.
Re-analyzing Past Dropout Flags for Accuracy, Precision, Sensitivity and Specificity

• Literature search.
• Queried multiple databases:
  – JSTOR, Google Scholar, EBSCO, Educational Full Text Wilson Web
• Studies were included that:
  – Were published since 1979
  – Examined High School dropout
  – Examined school-wide characteristics and included all students
  – A focus on the student level
  – Reported frequencies for re-analysis

• Initially yield 6,434 overlapping studies
  – 301 studies were read in full
  – 140 provided school-wide samples and quantifiable data
  – 36 articles provided enough data for accuracy re-calculations
  – Yield 110 separate dropout flags

• Relative Operating Characteristic (ROC) – *Hits versus False-Alerts*
Event table for calculating dropout contingency proportions

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Event</th>
<th>Dropout</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dropout</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
</tr>
<tr>
<td></td>
<td>True-positive (TP)</td>
<td>False-positive (FP)</td>
<td>Type I Error</td>
</tr>
<tr>
<td>Correct</td>
<td>a+c</td>
<td>b+d</td>
<td>a+b+c+d=N</td>
</tr>
<tr>
<td>Graduate</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
</tr>
<tr>
<td>False-negative (FN)</td>
<td>True-negative (TN)</td>
<td>Correct</td>
<td></td>
</tr>
<tr>
<td>Type II Error</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Precision \[= \frac{a}{a + b}\] \textit{Positive Predictive Value}

True-Positive Proportion \[= \frac{a}{a + c}\] \textit{Sensitivity} \textit{“Hits”}

True-Negative Proportion \[= \frac{d}{b + d}\] \textit{Specificity}

False-Positive Proportion \[= \frac{b}{b + d}\] \textit{1-Specificity} \textit{“False Alarms”}

An example of the true-positive proportion plotted against the false-positive proportion for Balfanz et al. (2007) comparing the relative operating characteristics (ROC) of each dropout flag.
Relative operating characteristics (ROC) of all dropout flags reviewed, plotted as the true-positive proportion against the false-positive proportion. Numbers refer to dropout indicator IDs.

Dropout Indicators Found for 1st Graders

By Sarah D. Sparks

As tracking data on students grow ever more extensive, some Maryland educators are finding that the early-warning signs of a student at risk of dropping out may become visible at the very start of their school careers.

Spotting Future Dropouts by Grade 1

In Montgomery County, Maryland, administrators have identified early-warning signs for students at risk of dropping out of school. They can identify flags for 75 percent of dropouts—by the second semester of first grade. Some of the warning signs are reading below grade level, absences of nine days or more, report card grades, incomplete assignments, and behavior. Montgomery County Evaluation Specialist Thomas (Chris) C. West, who created the formula for his district, commented in an Education Week article (07/29/2013), “You will not reduce dropout rates by identifying the students; it’s what you do with them. Early intervention strategy; they are part of an intervention strategy.

Can an algorithm ID high-school dropouts in first grade?

Early warning systems to detect high-school dropouts are all the rage in education data circles. See this post on a new early warning system in Wisconsin. Like the Wisconsin example, most data systems focus on identifying middle-school students. But what if researchers could use grades, attendance and behavior data to identify at-risk students as soon as possible — as early as first grade? That would really give counselors more time to try to motivate these kids and keep them in school!

Study: 1st Grade Dropout Indicators Found

news story by James Dugan | July 29, 2013

Related Topics: education, reform, dropout, first grade, graduation rates, school dropouts, rates, indicators, student achievement

While school districts and state departments of education have been tracking dropout indicators in students for a while, educators in Montgomery County in Maryland are building an early warning system designed to help students in the first grade.
Relative operating characteristics (ROC) of all dropout flags reviewed, plotted as the true-positive proportion against the false-positive proportion. Numbers refer to dropout indicator IDs.

Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin

Jared E. Knowles *
Wisconsin Department of Public Instruction

The state of Wisconsin has one of the highest four year graduation rates in the nation, but deep disparities among student subgroups remain. To address this the state has created the Wisconsin Dropout Early Warning System (DEWS), a predictive model of student dropout risk for students in grades six through nine. The Wisconsin DEWS is in use statewide and currently provides predictions on the likelihood of graduation for over 225,000 students. DEWS represents a novel statistical learning based approach to the challenge of assessing the risk of non-graduation for students and provides highly accurate predictions for students in the middle grades without expanding beyond mandated administrative data collections.

Similar dropout early warning systems are in place in many jurisdictions across the country. Prior research has shown that in many cases the indicators used by such systems do a poor job of balancing the trade off between correct classification of likely dropouts and false-alarm (Bowers et al., 2013). Building on this work, DEWS uses the receiver-operating characteristic (ROC) metric to identify the best possible set of statistical models for making predictions about individual students.

This paper describes the DEWS approach and the software behind it, which leverages the open source statistical language R (R Core Team, 2013). As a result DEWS is a flexible series of software modules that can adapt to new data, new algorithms, and new outcome variables to not only predict dropout, but also impute key predictors as well. The design and implementation of each of these modules is described in detail as well as the open-source R package, EWTools, that serves as the core of DEWS (Knowles, 2014).
Figure 1: ROC points for 110 published early warning indicators from Bowers & Sprott (2010). The diagonal line represents random chance, the top left corner represents perfect prediction.
http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM082
Models Tested:
Lasso
Subset Selection
Least Squares
Generalized Linear Models
Generalized Additive Models
KNN
Trees
Bagging boosting
Support Vector Machines
Ensembles

http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM082
Growth mixture model for the simultaneous estimation of latent trajectory classes using non-cumulative GPA from the first three semesters of high school.

Demographics:

Student:
- Female
- African American
- Asian
- Hispanic
- Non-Traditional Family
- SES

School:
- Urban
- Rural
- % Students Free Lunch

Behavior & Structure:

Student:
- Extracurricular
- Retained
- Negative Behavior

School:
- Student-Teacher Ratio
- Academic Press
- Small School
- Large School
- Extra-Large School

GPA 9S1 GPA 9S2 GPA 10S1
Intercepts Slopes
High School Dropout

Latent Trajectory Classes
C

Longitudinal Non-Cumulative GPA Trajectories in the First Three Semesters of High School

Mid-Decreasing: 10.8%
Low-Increasing: 13.8%
Mid-Achieving: 56.5%
High-Achieving: 18.9%

Dropout:

- Mid-decreasing & Low-increasing accounted for: 24.6% of the sample
- 91.1% of the dropouts

\[ n = 5,400 \]


Open Access: http://dx.doi.org/10.7916/D8NV9JQ0

Open Access: https://doi.org/10.7916/D8WM1QF6

Relative operating characteristics (ROC) of all dropout flags reviewed, plotted as the true-positive proportion against the false-positive proportion. Numbers refer to dropout indicator IDs.

Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes

Alex J. Bowers and Xiaoliang Zhou

Department of Organization and Leadership, Teachers College, Columbia University

ABSTRACT

Early Warning Systems (EWS) and Early Warning Indictors (EWI) have recently emerged as an attractive domain for states and school districts interested in predicting student outcomes using data that schools already collect with the intention to better time and tailor interventions. However, current diagnostic measures used across the domain do not consider the dual issues of sensitivity and specificity of predictors, key components for considering accuracy. We apply signal detection theory using Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) analysis adapted from the engineering and medical domains, and using the pROC package in R. Using nationally generalizable data from the Education Longitudinal Study of 2002 (ELS:2002) we provide examples of applying ROC accuracy analysis to a variety of predictors of student outcomes, such as dropping out of high school, college enrollment, and postsecondary STEM degrees and careers.

Journal Version: https://doi.org/10.1080/10824669.2018.1523734
Open Access Version: https://doi.org/10.7916/d8-nc5k-3m53
Online Supplement R Code: https://doi.org/10.7916/D8K94RDD
Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) to Evaluate the Accuracy of Predictors for College Enrollment and Postsecondary STEM Degree

<table>
<thead>
<tr>
<th>Predictor</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>0.767</td>
</tr>
<tr>
<td>Extracurricular activities</td>
<td>0.642</td>
</tr>
<tr>
<td>Extracurricular. activities (2004):</td>
<td>0.642</td>
</tr>
<tr>
<td>AP</td>
<td>0.565</td>
</tr>
<tr>
<td>Number of STEM Courses</td>
<td>0.957</td>
</tr>
<tr>
<td>Math t score (2002)</td>
<td>0.663</td>
</tr>
<tr>
<td>STEM course GPA</td>
<td>0.566</td>
</tr>
</tbody>
</table>
Contributions & Open Questions

For K-12 & Higher Ed Prediction Algorithms

• Visualize the data and pattern analytics
  – Can we see the individuals in the data and patterns?
  – Are there typology subgroups?

• Use accuracy comparison metrics and visualizations
  – ROC AUC, precision-recall, kappa

• Publish the code with example walk-throughs so that education data analysts can work with your algorithms.

• Do these accuracy metrics for predictors replicate across multiple district and state contexts?

• How to provide accurate predictors to teachers and leaders for their decision making?
Overview: Education Leadership Data Analytics (ELDA)

- AI in education, data analytics, and education data science. Hype versus reality.
  - What is Education Leadership Data Analytics (ELDA) and why should anyone care?

- Data science, and education leadership training

- What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?
  - Interaction and small group discussion

- Pattern visualization of student course interactions:

- Calculating predictor accuracy for early warning systems

Understanding How Educators Use Instructional Data Warehouses (IDWs) (or not) NSF collaboration:

- Building community and capacity in data use: Nassau BOCES and Teachers College collaborative

- Do teachers and administrators have different perceptions of data use in their schools and IDW use?
  - [https://www.tc.columbia.edu/articles/2020/january/how-education-data-can-help-teachers-raise-their-game/]
Building Community and Capacity for Data-Intensive Evidence-Based Decision Making in Schools and Districts - Teacher Data Use Survey (TDUS)

National Science Foundation
NSF DGE #1560720

Any opinions, findings, and conclusions or recommendations are those of the author and do not necessarily reflect the views of funding agencies
A Collaborative Partnership with Nassau County BOCES to Build Community and Capacity around Data Use and Evidence-Based Improvement Cycles

Nassau County Long Island New York:
- 56 School Districts
- Over 200,000 students
- 379 schools
Data use in schools:

- Evidence-based decision making in schools and districts is a growing area of research and practice.
- Teachers, school and district leaders have a wealth of knowledge about how they use data to inform their practice.
- Growing consensus in the research that training teachers and leaders how to use data to inform evidence-based improvement cycles can impact instructional improvement.

However:

- Much of this research does not take advantage of the complex and rich data-sources available to schools and districts.
- Teachers and administrators may disagree on what data is most accessible, useful and informative for instructional improvement.
- To date, the research has considered educator data use as a single dimension of high to low. Could there be subgroups of teacher and leader data use that may help inform professional development, evidence-based conversations and instructional improvement?
Purpose & Goals of the Overall Project

Purpose of the project:
To build a researcher-practitioner community around data-intensive evidence-based decision making through bringing together large and diverse school district datasets and then applying recent innovations from the data sciences to pilot and test data analytics with teachers and administrators.

Goals of this four year project:

1. Learn how data are used across districts and schools in Nassau County.

2. Understand how big data and data science techniques can help provide new opportunities for evidence-based improvement cycles.

3. Privilege the innovation in Nassau County, nationally, as a test-bed for innovation in evidence use to improve instruction.

4. Collaboratively build and test new analysis and data visualizations for schools using open-access code to address the issues educators see as most important, then release this code nationally to build a foundation for future innovations.
Survey:
Teachers and administrators asking them what they say they do with data, n=4,941.

- Four Statistically Significantly Different Types of Responders (LCA)

Qualitative Interviews:
40 educators from the survey. About 10 in each of the four LCA groups.

Today: Breakout Session 12:15-1:25: 
Measuring the Mindset of Educators
Elizabeth Young, Nassau BOCES
Sarah Weeks, Teachers College

Collaborative Workshop
Gather 50+ educators and 20+ education data scientists for a 2-day collaborative “datasprint” to co-design data visualizations that educators want using the data that matters to them.

Today: Breakout Session 12:25-1:25:
Working with Cognos 11 Visualizations
Jeff Davis, Nassau BOCES

Clickstream Logfiles:
All clicks in the Instructional Data Warehouse (IDW) > 100,000+ clicks

- Four Statistically Significantly Different Types of Responders (LCA)

Today: Breakout Session 1:35-2:35:
Clicks in Context
Robert Feihel, Nassau BOCES
Aaron Hawn, Teachers College & Penn

Today: Breakout Session 1:35-2:35:
Clicks in Context
Robert Feihel, Nassau BOCES
Aaron Hawn, Teachers College & Penn
TDUS: Teacher Data Use Survey

**2016**
- Grant awarded Preplanning
- TDUS into online form
- Meeting with pilot districts

**2017**
- Pilot TDUS
- Meeting with pilot districts
- Full administration of TDUS
- Data processing completed

**2018-2019**
- Teacher & Administrator Interviews 2018-2019 \( n=40 \)
- Data analysis of how the Instructional Data Warehouse (IDW) is used (or not)
- 2-Day workshop to pilot IDW tools and build new tools in collaboration with teachers and administrators
- Publish e-book, journal articles, R code datapipe, R code dashboards

**2020**
- Preplanning
- Meeting with pilot districts
- All districts invited to participate
- Data processing completed
Data Use Theory of Action

1. Access Collect
2. Organize Filter Analyze
3. Combine with understanding & expertise
4. Apply
5. Assess effectiveness

Response & Action

INFORMATION

KNOWLEDGE

DATA

OUTCOMES

Feedback

Data use theory of action - adapted from Ikemoto & Marsh (2007); Mandinach, Honey, Light, & Brunner (2008); Marsh (2012); Schildkamp & Kuiper (2010); Schildkamp, Poortman, & Handelzalts (2016)
Data use theory of action - adapted from Ikemoto & Marsh (2007); Mandinach, Honey, Light, & Brunner (2008); Marsh (2012); Schildkamp & Kuiper (2010); Schildkamp, Poortman, & Handelzalts (2016)
Purpose

The purpose of this study is to investigate the extent to which there are significantly different subgroups of responders to the Teacher Data Use Survey (TDUS) from across 56 school districts to examine a typology of school and district leaders, teachers and instructional support staff perceptions of data use.

Wayman, Wilkerson, Cho, Mandinach, & Supovitz (2016)
Research Questions

• To what extent are there significantly different subgroups of responders to the Teacher Data Use Survey?

• To what extent are teachers, school and district administrators, and instructional support staff’s perceptions on how teacher use data associated with membership in these subgroups of responders?

• Do these different types vary by how they click in the IDW dashboard system?

• What data do educators say are most accessible and useful to their practice?
Teacher Data Use Survey (TDUS)

• The TDUS is created and validated by the U.S. Department of Education Institute of Education Sciences, Wayman and colleagues (2016).

• The TDUS collects the perceptions of the data use by teachers in schools around issues of accessibility of data, frequency of data use, usefulness of data, actions with data, competence, attitudes, collaboration, organizational supports and leadership.

• **Survey Versions:** Teachers, Administrators (e.g. principals), Instructional support staff (e.g. instructional coaches)

• **Purpose:** Provide a survey instrument that enables district and school leaders to learn more about:
  – How teachers use data
  – Teachers attitude toward data
  – Teacher perception of supports for using data
  – How teacher collaborate to use data

• **Survey Administration:** The full survey administration was completed at the end of May 2017.
Figure 1. Teachers’ actions are at the center of the conceptual framework that describes their data use

- Competence in using data
- Attitudes toward data
- Collaboration
- Organizational supports

Actions
- Make decisions
- Evaluate problems
- Synthesize information
- Form questions
- Examine data
- Knowledge and practice

Student learning

Three different survey forms:

- School Leaders
- Teachers
- Support staff

Teacher Data Use Survey (TDUS):
- Online (emailed) survey link to all teachers, administrators (school and district) and instructional support staff for the entire county.
  - About 20,000 people
- Received 4,941 responses (25% response overall rate)
  - 3,783 full survey responses for dashboard reporting
  - Used multiple imputation for analytics (FIML)
- Some districts had as high as 65% response rate, many had less than 15% response rate, and a few less than 10% response rate.
  - Modal response rate by districts was 22%.
- By position:
  - 3,776 teachers
  - 405 administrators (building-level and district)
  - 760 instructional support staff

Results Reporting:
- TDUS dashboards generated overall, by district and by building
  - Dashboards generated in R. Open source code available by spring 2018.
- Meetings with superintendents, principals and teachers to discuss results
- Latent Class Analysis (LCA) (clustering) of the responses to ask to what extent is there one or more than one type of data user across these school districts?

Self-Reported Position Titles:

Teachers

Administrators
Purpose. The purpose of the Teacher Data Use Survey (TDUS) is to provide helpful information to district and school leaders about how teachers use data, their attitudes toward data, and the supports that help them use data.

Versions. A comprehensive perspective is possible because there are three versions of the TDUS—one for teachers, one for administrators, and one for instructional support staff (ISS).

Framework. The TDUS is based on a framework with five total components designed to measure the actions teachers take with data and then four supporting components that inform these actions: their competence in using data, their attitudes toward data, their collaboration, and their organizational supports.

Data Availability, Frequency of Use, and Usefulness. The TDUS begins with questions that collect descriptive information about various forms of data for the school or district. The first two pages show the availability of these data, the frequency of their use, and their perceived usefulness.

**Teacher Data Use Survey (TDUS) Results**

**Pilot Study Findings**

**August 6, 2017**

**IS THIS FORM OF DATA AVAILABLE TO TEACHERS?**

As part of the organizational support component, respondents share the availability of various forms of data.

**HOW FREQUENTLY DO TEACHERS USE THIS FORM OF DATA?**

As part of the action component, respondents share how frequently they use the forms of data to plan for instruction.

**HOW USEFUL IS THIS FORM OF DATA TO TEACHER PRACTICE?**

As part of the attitudes toward data component, TDUS asks how useful the various forms of data are to teacher practice.

**StateAssess** (e.g., NYS 3-8; NYSESLAT; NYS; Regents)

- Teachers
- Administrators
- Support Staff

**Benchmark** (e.g., START™; NWEA™; DIBELS®; AIMSweb®; Acuity®)

- Teachers
- Administrators
- Support Staff

---

**Table Representation**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Frequency</th>
<th>Very Useful</th>
<th>Useful</th>
<th>Somewhat Useful</th>
<th>Not Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>StateAssess</td>
<td>Teachers</td>
<td>11%</td>
<td>18%</td>
<td>28%</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td>Administrators</td>
<td>16%</td>
<td>29%</td>
<td>35%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Support Staff</td>
<td>9%</td>
<td>23%</td>
<td>41%</td>
<td>27%</td>
</tr>
<tr>
<td>Benchmark</td>
<td>Teachers</td>
<td>9%</td>
<td>22%</td>
<td>24%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td>Administrators</td>
<td>23%</td>
<td>37%</td>
<td>17%</td>
<td>24%</td>
</tr>
<tr>
<td></td>
<td>Support Staff</td>
<td>15%</td>
<td>38%</td>
<td>33%</td>
<td>15%</td>
</tr>
</tbody>
</table>
Survey Scales: The TDUS contains nine scales that measure either the actions teachers take with data or one of the four supporting components. These scales are in all three survey versions, although the question stems and/or question items may vary depending on the version.

Findings: The following section shares the findings for scales within each of the five components of the framework: actions, competence in using data, attitudes toward data, collaboration, and organizational support. The responses to all items are averaged for each respondent on a scale of 1 to 4, resulting in the respondent's scale mean for that component. Then, all scale means are combined by survey version.
Model of the Latent Class Analysis (LCA) of Responders to Teacher Data Use Survey (TDUS)

Teacher Data Use Survey Components (Indicators)
- Frequency of Data Use
- Data Usefulness
- Action with Data
- Competence
- Attitudes toward Data
- Collaboration
- Organization Support

Contextual Factors (Covariates)
- Teachers
- Administrators
- Instructional support staff

Latent Class C
Findings

### Latent Class Analysis Results and Fit Statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>Loglikelihood</th>
<th>LMR Test for k-1 classes</th>
<th>p</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two Classes</td>
<td>90330.101</td>
<td>90609.830</td>
<td>-45122.050</td>
<td>8941.043</td>
<td>0.000</td>
<td>0.780</td>
</tr>
<tr>
<td>Three Classes</td>
<td>88574.593</td>
<td>88997.439</td>
<td>-44222.296</td>
<td>1789.942</td>
<td>0.000</td>
<td>0.726</td>
</tr>
<tr>
<td><strong>Four Classes</strong></td>
<td><strong>87091.213</strong></td>
<td><strong>87657.176</strong></td>
<td><strong>-43458.606</strong></td>
<td><strong>1519.261</strong></td>
<td><strong>0.000</strong></td>
<td>0.707</td>
</tr>
<tr>
<td>Five Classes</td>
<td>86376.826</td>
<td>87085.906</td>
<td>-43079.413</td>
<td>754.356</td>
<td>0.401</td>
<td>0.742</td>
</tr>
<tr>
<td>Six Classes</td>
<td>85795.003</td>
<td>86647.201</td>
<td>-42766.502</td>
<td>622.496</td>
<td>0.067</td>
<td>0.777</td>
</tr>
<tr>
<td>Seven Classes</td>
<td>85311.265</td>
<td>86306.579</td>
<td>-42502.632</td>
<td>524.933</td>
<td>0.668</td>
<td>0.763</td>
</tr>
</tbody>
</table>

*Note: AIC = Akaike information criterion; BIC = Bayesian information criterion; LMR = Lo-Mendell-Rubin adjusted likelihood ratio test.*

*N=4941*
Four Significantly Different Subgroups of Responders to the TDUS: Latent Class Analysis (LCA)

Group 4 (24.2%)
High Data Use with Collaboration & Action

Group 2 (27.8%)
Data with low collaboration

Group 3 (25.5%)
Data Forward

Group 1 (22.5%)
Individual Classroom Focus
Instructional Data Warehouse Assessment Click Actions

Administrator Assessment Actions by LCA Data Use Type

n = 135 Administrator Users

Administrators n=135
Instructional Data Warehouse Assessment Click Actions

**Administrator Assessment Actions by LCA Data Use Type**

- **n = 135** Administrator Users
- Total Assessment Actions
- LCA Data Use Type
  - Class Focus
  - Data + Low-Collab
  - Data Forward
  - High Data + Collab

**Teacher Assessment Actions by LCA Data Use Types**

- **n = 147** Teacher Users
- Total Assessment Actions
- LCA Data Use Type
  - Class Focus
  - Data + Low-Collab
  - Data Forward
  - High Data + Collab

Administrators  $n=135$

Teachers  $n=147$
Academic Data Types Identified in Teacher Follow-up Interviews

Desired

Desired

Perception of Availability

Low/Low

High/Low

Low/High

High/High

Low/Low

High/Low

Low/High

High/High
Academic Data Types Identified in Teacher Follow-up Interviews

Perception of Availability

(low)  (high)

Commercial Reading & Math Assessments:
- i-Ready
- NWEA MAP Math & ELA
- Right Reasons Technology
- Imagine Learning
- ORI
- STAR
- Brigance
- Running Records
- Fountas & Pinnel
- DRA
- Dibbles
- Castle Learning
- AimsWeb

(1) Perception of Availability

- School/Grade Level Benchmark
- Regents Exam
- Teacher Observation
- Teacher Created Assessment

- Curriculum Provided Assessment

- NY State 3-8 Math & ELA Test

- International Baccalaureate Exams & Papers

- Current Grades

- Prior Grades
- AP Exams

- District or BOCES Benchmark

- Graduation
- Portfolio of Student Work

- Prior Teacher Anecdotal/Reflection - Acad

- ELL Teacher Data/ Note
- Special Education Testing/IEP

- NYSITELL/NYSESLAT

- SAT or ACT

- Homework

- Gross Motor/Fine Motor Development

(l)  (h)
Conclusion

• There are at least four different types of responders to the TDUS
  – The four subgroups value and collaborate around data in very different ways.
  – Dashboards plus the LCA subgroup analysis provides opportunities for potential targeted district capacity building and professional development around data use and evidence-based improvement cycles
  – For the highest responders on the Teacher Data Use Survey
    • Principals click the most while teachers click the least in the Instructional Data Warehouse (IDW).
  – The more available the data type, the more it is desired.

Find out more at the breakout sessions this afternoon!

• Measuring the Mindset of Educators: Attitudes and Perceptions of Data Use in Schools and Districts: 12:25-1:25 Albany room
  – Elizabeth Young & Sarah Weeks
• Working with Cognos 11 Visualizations: 12:25-1:25 Salon B
  – Jeff Davis
• Clicks in Context: Getting to Insight and Action from Data Warehouse Activity Logs: 1:35-2:35 Salon C
  – Robert Feihel & Aaron Hawn
Leading with Evidence in Schools: Data and Research Literacy

March 2 - 29, 2020

Upcoming course offerings:
Spring Session - March 2 - 29, 2020
Summer Session - July 6 - August 2, 2020
Fall Session - November 2 - 29, 2020

Where: Online asynchronous course

Registration Fee: $595

Group/Team Discount (5 or more):
25% off the registration fee, please contact cps@tc.columbia.edu to register

School administrators and educators employed by the New York City Department of Education are eligible for individual discounts of 25%. Please contact cps@tc.columbia.edu for more information.

Units Awarded:
Participants receive 20 Clock Hours and 20 CTLEs (applicable only to NYS residents).

http://tc.columbia.edu/cps/evidence
Overview: Education Leadership Data Analytics (ELDA)

• AI in education, data analytics, and education data science. Hype versus reality.
  • What is Education Leadership Data Analytics (ELDA) and why should anyone care?

• Data science, and education leadership training

• What components of a statewide longitudinal data system are needed for that system to best meet the respective needs of superintendent, principal and teacher leaders?
  • Interaction and small group discussion

• Pattern visualization of student course interactions:

• Calculating predictor accuracy for early warning systems

Understanding How Educators Use Instructional Data Warehouses (IDWs) (or not) NSF collaboration:

• Building community and capacity in data use: Nassau BOCES and Teachers College collaborative

• Do teachers and administrators have different perceptions of data use in their schools and IDW use?
Contributors to the Education Leadership Data Analytics Research Group:

April Bang  Aaron Hawn  Yilin Pan  Sarah Weeks  Yi Zhang
Kenny Graves  Elizabeth Monroe  Burcu Peckan  Yihan Zhao  Sen Zhang
Eric Ho  Xinyu Ni  Luronne Vaval  Seulgi Kang  Xiaoliang Zhou

Thank you!

Email: Bowers@tc.edu
Faculty Webpage: http://www.tc.columbia.edu/faculty/ab3764/
Research Group Website: http://www.tc.columbia.edu/elda/
Leading with Evidence in Schools Online PD Course: http://tc.columbia.edu/cps/evidence