Big Data & the Predictive Analytics Reporting (PAR) Project

Karen Swan, University of Illinois Springfield
Phil Ice, American Public University Systems
Rob Mitchell, American Public University Systems
Ben Arbaugh, University of Wisconsin Oshkosh
Peter Shea, University at Albany

“The calls for more accountability in higher education, the shrinking budgets that often force larger class sizes, and the pressures to increase degree-completion rates are all raising the stakes for colleges and universities today, especially with respect to the instructional enterprise. As resources shrink, teaching and learning is becoming the key point of accountability. “

-- Malcolm Brown & Veronica Diaz, 2011

“The rarely articulated implication of all of this data floating around is that un-augmented human cognition is no longer sufficient. Every day, every day, there is more to know, more ways to know it, and heightened expectations—by students, faculty members, alumni, football coaches, trustees, regulators, elected officials—that senior managers will do something efficacious with what they know. . . . The best tool in [their] battle against ignorance is advanced analytics.”

-- Thorton May, 2011

"PREDICTIVE ANALYTICS marries large data sets, statistical techniques, and predictive modeling. It could be thought of as the practice of mining institutional data to produce actionable intelligence."

-- CAMPBELL, DEBLOIS, & OBLINGER, 2010

"LEARNING ANALYTICS is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs."

-- 1ST INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 2011

Levels of Analysis

Exploratory Statistics

- Comprehensive
- Higher confidence level for prediction
- 1% of solutions

Inferential Statistics

- Single system
- Low confidence levels for prediction
- 9% of solutions

Descriptive Statistics

- Single system
- Subjective interpretation
- 90% of solutions

The Predictive Analytics Reporting Framework

Phil Ice, Ed.D.
VP, Research & Development
American Public University System

Predictive Analytics Reporting

- WCET initiative funded by Bill and Melinda Gates Foundation
- Multi-Institutional aggregation and analysis
- Began in May, 2011

What We Found

- Taking concurrent courses, early in their program, significantly reduces that chances that at-risk students will be successful
- For all other students there is an incredible degree of complexity in predicting success

What We Want to Achieve

- Demonstrate that this CAN be done
- Demonstrate the the methodology is scalable
- Development of future directions (variables) to be included in research
- Completion of all goals on or before (preferable) the end of the project

Why PAR Matters

- retention in face-to-face courses is higher than in online courses, by anywhere from 2-5% (Moore & Fetzner, 2009) to 50% (Dutton, Dutton & Perry, 1999)

No Common Measures

- course completion (% of students still enrolled at end of course, still enrolled and not failing) but date of census matters
- success (% of students w/o W, F, D (or C in graduate courses)
- retention (semester to semester enrollment)
- progression to degree/degree completion

Factors Affecting PAR Students

- learner characteristics
- course characteristics
- support characteristics

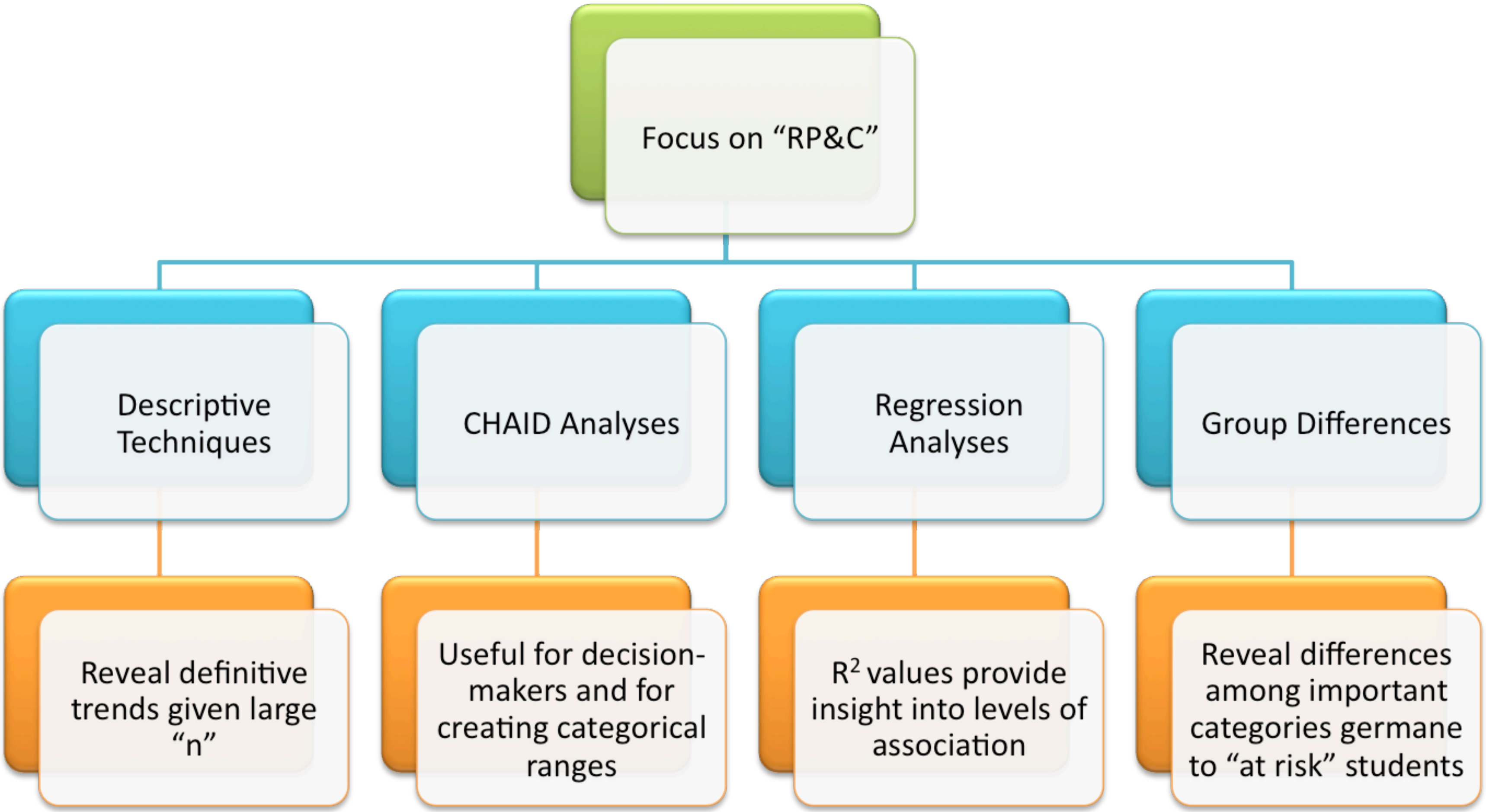
Operational Issues

- Analytical/Methodological
  - Operationalizing progression, retention and completion varies by institution.
- Organizational
  - Academic semesters/periods vary greatly by institution, including whether or not multiple courses pursued simultaneously.

Operational Issues

- Pedagogical
  - Developmental Education courses play varying roles at different institutions. These courses, often not-for-credit, have a profound impact on measures of RP&C.
- Technological
  - Workflow involving multiple institutions' data and FERPA compliance is a complicated process.

Analysis Protocols



Bouldérado Variables

Institution Identifier	Total Course Extensions	Gender	Veteran
PARStudentID	Total Degree Extensions	Non ResAlien Status	Transfer Credits
DegreeType	Previous Term Mean GPA	Race	Program Changes
AcademicLevel	Prior Term Withdrawals	Ethnicity	Prior Degree Completions
CIPCode	Degree Hours Attempted	Course Start Date	CourseGrade
MultipleMajor	Degree Hours Completed	Course End Date	DevEd Courses Attempted
Academic Status	Course Size	DOB	DevEd Courses Completed
Institution Student Course Completes	Concurrent Credit Bearing Courses	Military Classification	Degree Start Date

Interpreting Regression

- ***Dependent Variable***: Grade Point Average
- ***Independent Variables***: Ball of Wax
- Results for 5 most influential variables when examining only those students who had >5 Degree Hours Completed:

Variable		Standardized Coefficients	R Square Change
Transfer Credits	☺	.150	2.20%
Race - White	☺	.089	1.29%
Prior Term Withdrawals	☹	-.080	0.68%
Prior Degree Completions	☺	.076	0.57%
Race - Black	☹	-.050	0.21%

Interpreting Regression

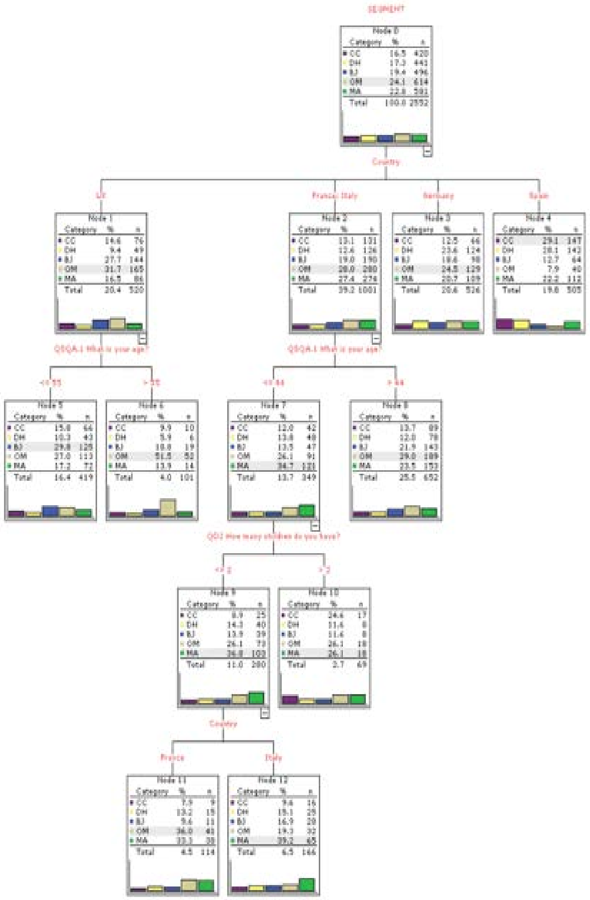
- Dependent Variable: Grade Point Average
- Independent Variables: Ball of Wax
- Results for 5 most influential variables when examining only those students who had <= 5 Degree Hours Completed:

Variable		Standardized Coefficients	R Square Change
Concurrent Courses	☹	-.419	26.70%
Associate's Program	☹	-.132	3.70%
Prior Degree Completions	☺	.153	2.30%
race	☹	-.095	0.80%
Transfer Credits	☺	.105	0.90%

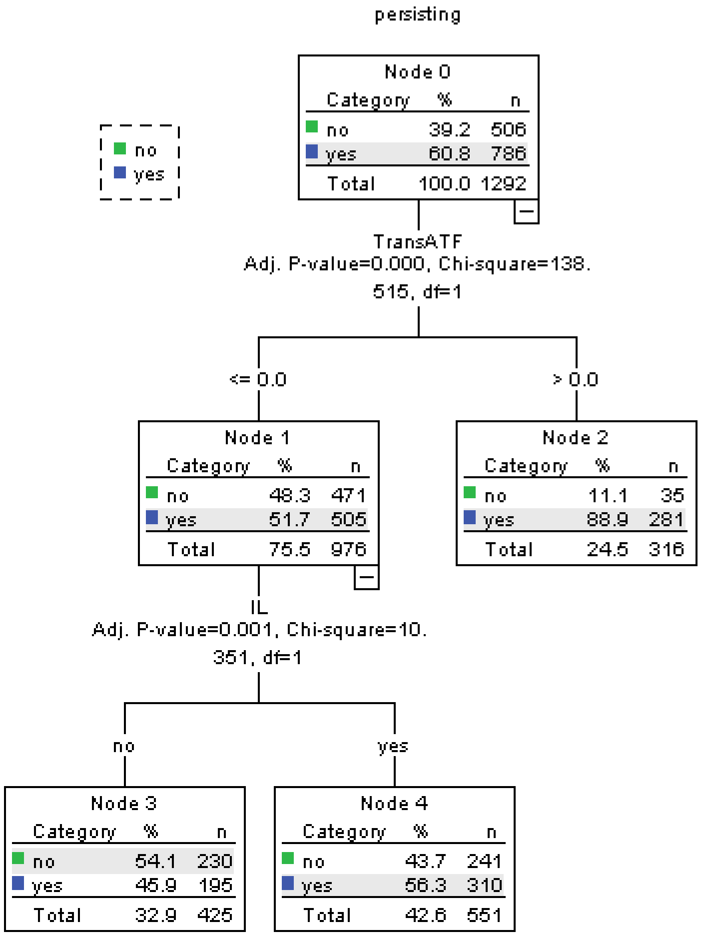
Essentially Infinite Patterns

- One small segment from one institution
- No two were alike

Example



- transferring credits into online programs after initial enrollment was a very strong predictor of persistence
- almost 90% of the students who did so persisted



UIS example

Technical Issues

Rob Mitchell
American Public University Systems

Privacy Issues

- FERPA considerations

Preparing the Technology

Speaking the Same Language

Potential Implications of PAR for Institutional and Online Teaching and Learning Research

J. B. Arbaugh
University of Wisconsin Oshkosh

Institutional Implications?

- Incentives for increased institutional collaboration and consortia?
- A complement, competitor, or successor to organizations such as Sloan-C and/or Quality Matters?
- Will the potential higher data accessibility elevate the profile of online courses in the pursuit of higher completion rates?

Research Implications?

- Increased heat on OTL researchers to address generalizability of findings
- Could subsequent PAR initiatives lead to the “Walmartization” of OTL research?
- How might OTL researchers respond?
  - Customization – targeted questions to targeted audiences (i. e. DSJIE)
  - Specialization – focus on aspects of online offerings beyond scope of PAR (i.e. graduate education)
  - Can’t beat ‘em, join ‘em – Seek access to PAR data through institutional or individual membership

Promises and Pitfalls of PAR vis a vis Big Data

Peter Shea, External Evaluator
University at Albany, SUNY

Promises/Possibilities

- Promise is to help, to identify problems so they can be addressed
- Struggling students everywhere can benefit
- Cost can be saved
- Efficiencies created
- Hard questions answered
- Citizens more efficiently prepared to be productive members of society

Concerns/Fears

- Certain "joylessness" implied by prediction
- Do we risk our humanity with the "end of theory" (as Chris Anderson puts it)?

Anderson: In the era of big data:

- "Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."

http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Danah Boyd

- Do we lean on the numbers too much? What is gained? What is lost?
  - (Hint: Maybe privacy, security, intellectual property)

Danah Boyd

- 1) Bigger Data are Not Always Better Data
- 2) Not All Data are Created Equal
- 3) What and Why are Different Questions
- 4) Be Careful of Your Interpretations
- 5) Just Because It is Accessible Doesn't Mean Using It is Ethical

http://www.danah.org/papers/talks/2010/WWW2010.html

Mckinsey

- Big problem is lack of access to the “right” big data, ability to use it well, and incentives to get it…

http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

McKinsey

- "Organizations need not only to put the right talent and technology in place but also structure workflows and incentives to optimize the use of big data."
- "Access to data is critical—companies will increasingly need to integrate information from multiple data sources, often from third parties, and the incentives have to be in place to enable this."
  - Replace "organizations and companies" with "universities" *then* "incentivize"…

Factors known to influence college persistence: *Access this data*…

- **Psychological** variables such as student motivation, goal commitment, and **coping strategies**; academic self-confidence, self efficacy
- **Environmental**/external variables such as family responsibilities, **family support**, language ability.
- **Situational** - learner life changes, i.e. alterations in lifestyle or circumstances that interfere with the expectations, commitment, and goals of the individual student. Examples include **personal illness**, changes in job status, and **family problems**, insufficient time, unexpected events.

Factors known to influence college persistence: *Access this data*…

- **Institutional** experiences such as co-curricular involvement, **academic integration** , quality of course instruction and **quality of interaction with instructor**; social integration such as quality of interactions with peers, and/or staff; academic guidance from advisors , technical support services.
- **Andragogical** - self-directed learning, pedagogical dimension influenced by the **freedom** to which the learner is given **to set learning goals**, to identify and use resources, to determine the effort and time to be allocated to learning, and to **decide how and what kind of evaluation** of the learning will take place.

arbaugh@uwosh.edu
rmitchell@apus.edu
pice@apus.edu
pshea@albany.edu
kswan4@uis.edu

THANK YOU