The Predictive Analytics Reporting Framework Proof of Concept

Ellen Wagner, WCET
Pearl Iboshi, University of Hawai‘i System
Hae Okimoto, University of Hawai‘i System
November 1, 2011
The Predictive Analytics Reporting (PAR) Framework

- Large N statistical tests and analyses to identify factors affecting retention, progression and completion.
- Focused on federating and aggregating data from more than 640,000 de-identified student records and 3 million course records contributed by six institutional partners.
- Moving toward Phase 2, where we are adding another 12 – 14 institutions to add another 2 million student records to the master data set.
PAR Framework Questions

- What factors influence student loss/retention and momentum/completion?
- How do the factors affecting loss differ from indicators of completion?
- Are there unique demographic, pedagogical or institutional factors affecting loss/retention and momentum/completion?
PAR Institutional Partners

- WCET, Managing Partner
- American Public University System
- Colorado Community College System
- Rio Salado College
- University of Hawaii System
- University of Illinois – Springfield
- University of Phoenix
PAR Framework Process Model

1. Identify variables common across institutions
2. Federate de-identified records of online students
3. Aggregate all data into a single pool
4. Apply exploratory statistical tests
5. Apply methods to support pattern analysis
6. Reporting and KM accelerate next round
### Variables Defined at June “Boulderado Summit”

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

*First Looks: Early Findings Shared at WCET*
PAR Dataset - Before & After

Before (June 7, 2011)
- 29 variables
- Agreement on definition

After (Sept 6, 2011)
- 33 variables
- 40 codebook updates
- 600+ email exchanges
- 1 day-long data meeting
- 12 weekly meeting discussions
Focus on “RP&C”

Descriptive Techniques
- Reveal definitive trends given large “n”

CHAID Analyses
- Useful for decision-makers and for creating categorical ranges

Regression Analyses
- $R^2$ values provide insight into levels of association

Group Differences
- Reveal differences among important categories germane to “at risk” students

* RP& C = Retention, Progress and Completion
View from an Institutional Partner Perspective
THANK YOU