Overview of goals and pitfalls of data science and education

KARAN WATSON, SR. PROFESSOR ECEN, CO-DIRECTOR IIEI
How has data science been used in higher education

- Recruiting
- Retention
- Progress
- Course Success
- Concept Mastery
- Improvement of Instruction
- Improvement of Programs
- Elimination of Programs
Survey of Institutions in 2018 (by AIR, NASPA, EDUCAUSE)

- Report by Parnell, Jones, Wesaw, Brooks with over 750 institutions responding
  - Units that are using data are very siloed
    - 96% working to improve student outcomes (mainly retention and course success)
    - 71% working to improve efficiencies of programs or services
    - 39% looking to eliminate or reduce programs
  - Most universities are not tracking benefits vs. costs
  - Most are focused on first year students
  - 76% using for descriptive behavior, 62% for predictive behaviors
### Table 1. Institutions' Investment in Data and Analytics, by Institution Size

<table>
<thead>
<tr>
<th></th>
<th>Descriptive</th>
<th>Predictive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under 1,000 (N = 99)</td>
<td>1,000–4,999 (N = 336)</td>
</tr>
<tr>
<td>No investment</td>
<td>12%</td>
<td>7%</td>
</tr>
<tr>
<td>Minor investment</td>
<td>51%</td>
<td>57%</td>
</tr>
<tr>
<td>Major investment</td>
<td>34%</td>
<td>32%</td>
</tr>
<tr>
<td>Don't know</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note.* Percentages may not total to 100% due to rounding.
Table 3. Focus of Studies in Support of Student Success (N varies)

<table>
<thead>
<tr>
<th></th>
<th>First-year students</th>
<th>Sophomores</th>
<th>Transfer-in students</th>
<th>Student athletes</th>
<th>Students of color</th>
<th>LGBTQIA students</th>
<th>Nontraditional students</th>
<th>First-generation students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student pipeline</td>
<td>85%</td>
<td>28%</td>
<td>58%</td>
<td>27%</td>
<td>48%</td>
<td>5%</td>
<td>33%</td>
<td>44%</td>
</tr>
<tr>
<td>Academic progress and success</td>
<td>82%</td>
<td>53%</td>
<td>56%</td>
<td>39%</td>
<td>58%</td>
<td>9%</td>
<td>39%</td>
<td>54%</td>
</tr>
<tr>
<td>Efficiency of degree completion</td>
<td>68%</td>
<td>45%</td>
<td>38%</td>
<td>19%</td>
<td>35%</td>
<td>7%</td>
<td>27%</td>
<td>32%</td>
</tr>
<tr>
<td>Career pathways and postgraduation outcomes</td>
<td>56%</td>
<td>32%</td>
<td>45%</td>
<td>21%</td>
<td>41%</td>
<td>7%</td>
<td>30%</td>
<td>39%</td>
</tr>
<tr>
<td>Student ability to afford higher education</td>
<td>48%</td>
<td>29%</td>
<td>30%</td>
<td>12%</td>
<td>23%</td>
<td>5%</td>
<td>22%</td>
<td>28%</td>
</tr>
</tbody>
</table>
### Table 4. Types of Studies in Support of Student Success (N varies)

<table>
<thead>
<tr>
<th>Study Area</th>
<th>Institution is not conducting these studies</th>
<th>Institution is planning to conduct these studies within the next year</th>
<th>Institution is conducting these studies but not annually</th>
<th>Institution conducts these studies annually</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career pathways and postgraduation outcomes</td>
<td>8%</td>
<td>11%</td>
<td>18%</td>
<td>63%</td>
</tr>
<tr>
<td>Student pipeline</td>
<td>2%</td>
<td>11%</td>
<td>32%</td>
<td>54%</td>
</tr>
<tr>
<td>Graduate student progress</td>
<td>35%</td>
<td>6%</td>
<td>17%</td>
<td>42%</td>
</tr>
<tr>
<td>Faculty workload and performance</td>
<td>21%</td>
<td>10%</td>
<td>28%</td>
<td>40%</td>
</tr>
<tr>
<td>Academic progress and success</td>
<td>6%</td>
<td>20%</td>
<td>41%</td>
<td>33%</td>
</tr>
<tr>
<td>Student ability to afford higher education</td>
<td>46%</td>
<td>22%</td>
<td>21%</td>
<td>11%</td>
</tr>
<tr>
<td>Efficiency of degree completion</td>
<td>16%</td>
<td>32%</td>
<td>42%</td>
<td>10%</td>
</tr>
</tbody>
</table>
Figure 4. Institutions' Use of a Vendor for Student Success Studies

- **Descriptive studies (N = 164)**
  - Yes: 50%
  - No: 43%
  - Don't know: 7%

- **Predictive studies (N = 165)**
  - Yes: 72%
  - No: 24%
  - Don't know: 4%

- **Early-alert system (N = 167)**
  - Yes: 51%
  - No: 46%
  - Don't know: 4%
Figure 6. Agreement With Statements on Data and Analytics (N = 331–432)

- In conducting student success studies, privacy rights are respected. 94% strongly agree or agree, 5% strongly disagree or disagree.
- To stay competitive, we must continue to invest in student success analytics. 88% strongly agree or agree, 8% agree, 4% disagree.
- The data used for student success analytics are accurate. 79% strongly agree or agree, 13% agree, 7% disagree.
- The results of analytics studies are used properly; wrong conclusions are not drawn. 54% strongly agree or agree, 28% agree, 18% disagree.
- We are able to implement the results of student success analytic studies effectively. 40% strongly agree or agree, 39% agree, 21% disagree.
- I am concerned that my institution depends on the quality of vendor algorithms that we do not fully understand. 24% strongly agree or agree, 12% agree, 64% disagree.
- I am concerned that my institution relies on blackbox algorithms to inform decisions. 12% strongly agree or agree, 20% agree, 68% disagree.
At TAMU

- Using ten years of historical data, EAB provides a risk level to currently enrolled undergraduates as to their likelihood to graduate from TAMU. There are three risk levels: High, Moderate and Low. The model looks at both pre-college characteristics (ACT/SAT, ethnicity, first-generation status, high school class rank, etc.) and performance at TAMU (hours completed, hours completed v attempted, GPA, etc.). Currently rerunning on 4 years instead of 10 years of data.

- The product provides a one-stop shop for advising reports. If an advisor asks students to seek other campus resources, advisors can see if students complied with the request v emails and phone calls to campus contacts.

- Students are able to schedule appointments with their advisors through the platform, as well as through the student app that we also purchased with the platform.

- Students receive automatic reminders of appointments.

- Advisors are able to create watch lists of students based on GPA, hours, probation status, classification, etc. and push campus resources or appointment requests.

- EAB provides a list of courses that are most predictive of success in a given major and suggested grades in such courses.

- The app alerts students to blocks on their pre-registration at the time they occur.

- There is an Early Alert System within EAB that I have found to be cumbersome for our faculty, in addition to the limited access we give faculty to student information. Therefore, we are working with Instructional Technology Services on campus to provide reports on students earning D’s or less in the 15 highest enrollment/highest Q/D/F courses.
To go to Concept Mastery or Instructional improvement

- Instructors must get involved in designing assessments and inputing data
- Instructors must understand learning and appropriate lag times in student behavior so that the system does not over react
- Students must participate to get quality data

- Is it worth it? Only if the data is high quality