Executive Summary

This report offers recommendations regarding the use of learning analytics (LA) data and tools to enhance teaching and learning at Colorado State University. Members of the task force focused on

- institutional goals for using LA tools at CSU;
- faculty, student, advisor, researcher, and administrator goals for using LA tools;
- the types of LA tools that have been and might be used at CSU to pursue institutional, programmatic, and individual goals;
- ethical issues informing the collection of LA data and the use of LA tools;
- professional development initiatives that would enhance the understanding and use of LA tools;
- directions for coordinating the development, use, and oversight of LA data and tools.

The task force makes the following recommendations, each of which is discussed in detail later in this report.

Learning Analytics Purposes and Infrastructure

1. Learning analytics data and tools should be used to enhance student learning and success.
2. The existing and emerging LA dashboards in Canvas should be promoted as a default LA environment for all Canvas courses, while more advanced tools should be promoted for use in selected instructional areas where the impact on student success will be greatest (see recommendation 15 below).
3. CSU should continue to use EAB Navigate to support our student success efforts. This tool is curricular in nature, focusing on larger patterns (above the level of individual courses) that can inform advising. It also offers some student-centric predictive analytics on success in majors that could be used in a limited way at CSU.
4. CSU should use LA data and tools to support and automate Early Performance Feedback efforts.
5. CSU should use automated messages (nudges and alerts) sent through LA tools to encourage student learning and success.
6. CSU should continue to explore the efficacy and ethics of using zero-day predictive analytics in courses.

Ethical and Informed Use of Learning Analytics Data and Tools

7. CSU should recognize its ethical obligation to use LA data and tools to enhance student learning and success
8. CSU should commit to using LA data and tools in ways that adhere to principles outlined in the Ethical Principles of Learning Analytics document.
9. CSU should avoid using LA data as the sole or primary source of evidence of teaching or advising effectiveness.

10. CSU should require LA tools and data solutions providers to conform to CSU policies on use of sensitive LA data.

11. CSU should publicize to faculty and staff its guidelines for working with vendors who collect data on students and instructors.

**Pedagogy and Best Practices**

12. With predictive analytics, instructors and advisors should focus on delivering positive messages that will promote student success, such that even if the predictions are not completely accurate, any ensuing actions only improve student learning and performance.

13. CSU should provide faculty and staff with guidance in understanding and using LA tools to promote student learning and success, particularly where sensitive Personal Identifying Information (PII) is exposed. Instructors using visible PII should be required to complete a course in ethical use and behaviors associated with the exposure of such information.

14. Effective use of LA tools is predicated on the availability of meaningful data. To provide that data, CSU faculty should be trained in appropriate use of the Learning Management System and in related learning tools.

15. Detailed predictive LA should be targeted toward areas where it will be most beneficial, especially in lower-division courses with high DFW grade rates. A specific initiative should be developed in this regard, including departmental and faculty engagement (typically course coordinators), required training for instructors, and support for deployment when all conditions have been met. TILT should develop and lead this initiative.

**Planning and Reporting**

16. CSU should continue the Learning Analytics Steering Committee for at least three more years.

17. CSU should develop a strategic plan to align LA efforts with our mission and student success initiatives.

18. CSU should conduct a retreat to develop a LA implementation plan.

19. CSU should document how we are using LA data and tools to enhance teaching, learning, and student success.

**Research**

20. CSU’s use of LA data and tools should be informed by research. And that research should be inform best teaching and learning practices, subsequent research, and the existing LA knowledge-base.

The task force does not recommend additional expenditures beyond what is currently budgeted for learning analytics activities. However, several of the initiatives discussed in this report will require funding down the road and some of these costs are likely to become recurring expenses. One element requiring further definition is how students, faculty, advisors and resident assistants will use LA data, who will interact with the students under what circumstances (including workflows and hand-offs), and which LA elements are to be automated by machine algorithm.
1. Introduction

This report offers recommendations from the 2018-19 Task Force on Learning Analytics (LA) at Colorado State University. Following overviews of the Unizin consortium and then of learning analytics data and tools, the report focuses on four primary issues:

- LA tools that are or might be used at CSU to enhance teaching effectiveness, student learning, and student success.
- Ethical and privacy concerns related to the use of LA data and tools at CSU.
- Professional development for users of LA data and tools.
- Coordination of LA efforts at CSU.

This report also recommends investments to support these initiatives and includes several appendices.1

2. Unizin and Its Contributions to the Use of Learning Analytics at CSU

In 2014, a consortium of eleven distinguished higher education institutions came together with several goals:

- to regain some control over the digital learning ecosystem to encourage vendor development in directions that would be more beneficial to higher education,
- to share the costs of development and operations in two areas of instruction: sharing learning objects and sophisticated learning analytics tools and platforms, and
- to take advantage of consortial purchases for a much larger student population thus achieving significant economy of scale.

CSU participated in the planning of Unizin (https://unizin.org) and was among the first four founding members of the consortium. As a result of joining Unizin, CSU transitioned from the BlackBoard LMS to Canvas. This has been a strongly positive experience for faculty and students at CSU, as Canvas is much easier to use, open in its architecture, and continuously upgraded.

### Acronyms Used in this Report

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programming Interface, a technology used for accessing data in IT systems</td>
</tr>
<tr>
<td>ASC</td>
<td>Academic Success (or Support) Coordinator</td>
</tr>
<tr>
<td>BNED</td>
<td>Barnes and Noble Education, which supplies Unizin members with LoudSight (LS)</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Values, a file in one of the Excel formats</td>
</tr>
<tr>
<td>EPF</td>
<td>Early Performance Feedback, an initiative currently under the auspices of the Collaborative for Student Achievement</td>
</tr>
<tr>
<td>IS</td>
<td>Information Systems</td>
</tr>
<tr>
<td>LA</td>
<td>Learning Analytics</td>
</tr>
<tr>
<td>LMS</td>
<td>Learning Management System, in our case Canvas from the company Instructure</td>
</tr>
<tr>
<td>LS</td>
<td>LoudSight, Unizin’s Learning Analytics platform, supplied by BNED</td>
</tr>
<tr>
<td>PII</td>
<td>Personally Identifiable Information</td>
</tr>
<tr>
<td>SIS</td>
<td>Student Information System, in our case Banner/ARIES</td>
</tr>
<tr>
<td>UDP</td>
<td>the Unizin Data Platform that holds SIS and Canvas data for Unizin members</td>
</tr>
<tr>
<td>Zero-day</td>
<td>Predictive Learning Analytics of student success in the course, produced before the class begins</td>
</tr>
</tbody>
</table>

---

1 Some passages in this report are drawn from a forthcoming article about learning analytics by Mike Palmquist (2019). These passages are used with permission.
However, the primary reason CSU joined Unizin was to advance our understanding and use of learning analytics data and tools at the level of individual student. Our desire was and continues to be engaging in what has been called “Educational Moneyball,” using big data from learning management systems (in our case, Canvas) in combination with data from student information systems (in our case, Banner) to produce learning analytics predictions for students within individual course sections. Through collaboration across the Unizin consortium, systems and services can be developed and implemented for all Unizin member institutions, under a “divide by N” cost model and the power of multiple perspectives on challenging problems, rather than relying on less capable systems developed and deployed by individual institutions.

A key benefit of our membership in Unizin has been the opportunity to work with some of the best institutions in the country to advance the state of the art in learning analytics and bring the benefits of using learning analytics to our students, faculty, and staff. A central element in our LA environment is the Unizin Data Platform (UDP), which holds both in-course data from Canvas and SIS data at the individual student level. Additional information pertinent to the Unizin teaching and learning ecosystem is provided below in this document.

Recently, additional institutions have joined Unizin, and still more are interested in doing so as well. As more institutions join the consortium, its ability to develop and implement sophisticated learning analytics platforms, tools, and data repositories will continue to grow.

3. **An Overview of Learning Analytics**

Learning analytics has been a growing area of discussion and concern among scholars for more than a decade (Daniel, 2014; Fournier, Kop, & Sitlia, 2011; Retalis et al., 2006; Siemens & Long, 2011; Viberg et al., 2018). Certainly, using data to attempt to understand student learning performance is a long-standing practice in educational research. Recent improvements in our ability to assemble and analyze large sets of data before, during, and after a course has been offered, however, have enhanced our ability to understand how various learning behaviors, instructional practices, demographics, academic history, and instructional materials shape student learning, faculty teaching effectiveness, and student retention and success.

Instructors typically use LA data and tools to

- gain a sense of the likelihood of success of students prior to the start of a course,
- identify students who may be in danger of failing or performing poorly in a course while the course is being offered,
- identify learning behaviors that are correlated with student success (or the lack thereof) in a course,
- identify course materials and assignments that are correlated with student success, and
- encourage students to engage in effective learning behaviors and engage with relevant instructional materials (e.g., through messaging).
Sources of LA data include:

- learning management systems, which provide information about logins, completion of assignments and homework, course materials accessed over time, performance on quizzes and exams, access to files, and use of discussion forums, among other data (Daniel, 2015; Zhang et al., 2018)
- learning tools provided by vendors and publishers, such as adaptive learning tools and interactive exercises (Lewkow et al., 2015) as well as learning platforms (including McGraw-Hill’s Connect and Macmillan’s Achieve) that provide information about student behaviors and performance, typically with the goal of identifying students who might benefit from intervention by the instructor
- eReaders (such as Unizin’s Engage platform), video players, and other tools for accessing and interacting with course content (Junco & Clem, 2015; Shoufan, 2018)
- “multimodal” data sources, which can reveal student location and other activities in real time, such as posting to social media and accessing wireless networks, by drawing on data from the Internet of Things, cloud data storage, and wearable technologies (Di Mitri et al., 2018)
- written texts produced in formal and informal assignments, including journaling and posts on discussion forums ( McNely et al., 2012; Shum et al., 2016; Wise, Zhao, & Hausknect, 2013; Yu et al., 2017)

These data are often analyzed in combination with academic history, such as scores on college entrance examinations and performance in high schools—for example, high school GPA, CSU GPA, performance on related courses—as well as demographic information drawn from a student information system, such as race, ethnicity, gender, first-generation status, and financial aid information. In some cases, LA data from a specific course will be analyzed in combination with data about student participation in institutionally-supported activities, such as attending tutoring and study group sessions and meeting with faculty and academic advisors.

The functions offered by LA tools typically include:

- **Dashboards and reports.** Some LA tools offer dashboards that provide instructors and advisors with at-a-glance information on students’ behaviors and success to date in a course, potentially enabling instructors and/or advisors to better guide individual students and better adjust instruction to an entire class or to subsets of students. Many LA tools allow for customization of reports, allowing faculty and advisors to focus on specific students or groups of students or on particular behaviors or performance metrics. Some LA tools also offer dashboards that students can use to view their progress in a course, examine their own behaviors, set goals, and, in some cases, obtain advice about study behaviors that might increase their likelihood of completing the course successfully.

- **Messaging.** A number of LA tools can provide students with instructor-, advisor-, or auto-generated-communication to prompt stronger study behaviors. They can also provide instructors and advisors with information about student behaviors. Automatic messages sent to students are often referred to as nudges, while those which are also sent to instructors and advisors are typically referred to as alerts. Depending on the capabilities of the system that
generates them, nudges can be customized in advance or modified on the fly by instructors. When certain conditions are met, such as a set period of time following the last login to the system, failure to complete a quiz by a deadline, or failure to submit homework by a due date, **nudges** are sent to students. Depending on the settings in the system, alerts might also be sent to instructors and/or advisors. Messages may include updates on performance, suggestions about effective learning behaviors, information on associations between students’ uses of such behaviors and academic achievement, or other content. In some cases, alerts require a student to meet with the instructor or advisor in order to “clear” the alert.

A key issue associated with messaging involves how messages are constructed. On one level, it is a rhetorical issue, where the focus is on how to design messages that influence students in desirable ways. For example, designers of nudges and alerts seek to understand which rhetorical strategies are most likely to lead students to engage in behaviors that enhance their success in a course. On another level, it is a teaching and learning issue, where the focus is on which behaviors to address and which to encourage. For example, designers of nudges and alerts would benefit from understanding which behaviors are likely to lead to improved student learning and at what points it would be optimal to send messages. In the long run, the two levels intersect, in that the more knowledge is gained about which behaviors support learning, the more this information can be leveraged to increase the rhetorical effectiveness of nudges and alerts.

- **Predictions of Student Success.** Some LA tools offer dashboards and reports that provide instructors and advisors with predictions of student success at the level of individual students or entire classes. For ease of interpretation, these predictions are often communicated as scores on a specified scale. These scores are based on calculated probabilities of student success (e.g., earning a grade of A, B, C) or failure (e.g., earning a grade of D, F, W) in a course or program. Prior to the start of a course, such scores are based solely on academic history and demographic information. As the course unfolds, these scores are updated to reflect student behaviors and performance in the course. Typically referred to as “predictive analytics” or “predictive learning analytics,” these scores are derived through algorithms that operate on available data. On an individual level, these scores are subject to statistical variation, and are thus certain not to be completely accurate. However, improved accuracy results when these predictions are used in combination with what instructors learn from students through personal interaction and observation. Viewed in the aggregate, for example over all students in a course section, predictive analytics scores can provide an indication of the overall progress of students enrolled in a course.

In addition, a related set of tools—such as EAB’s Navigate—are being used to help institutions identify courses in which students struggle and, perhaps more importantly, to reveal course combinations within an academic term or course sequences across academic terms that appear to be correlated with lack of success.
Most LA platforms are produced by external vendors, whose policies differ regarding student data privacy, data stewardship, and the sharing of proprietary algorithms that govern dashboard scores. The use of these tools carries the potential risk that sensitive student data might be used in ways that are not aligned with CSU’s commitment to protect students’ information. That commitment is articulated in a recent set of recommendations found in *Ethical Principles of Learning Analytics at Colorado State University* (see Appendix A), which was produced by a Task Force charged by the Faculty Council Committee on Teaching and Learning. These recommendations provide the foundation for the suggestions regarding ethical use of LA data and tools in this document. Examples of misaligned data uses include vendors selling students’ class performance data to prospective employers or other third parties and using these data to develop additional tools then sold to CSU and other institutions. A benefit of our membership in Unizin is that Unizin is wholly owned and operated by its member institution. Since Unizin is not an external vendor, our use of Unizin LA tools obviates most of the ethical and privacy issues that might be encountered with use of products provided by external vendors.

Despite the risks discussed above, use of tools that generate LA data (and of the data themselves) to promote learning and student success is increasingly common (Viberga et al., 2018). As the major publishing houses reconfigure their products, moving away from traditional textbooks to online learning tools, it is certain that this trend will intensify.

4. **LA Tools in Use at CSU**

Several LA tools and platforms have been piloted or are in use at CSU, and several others are being considered for use.

**EAB NAVIGATE**

EAB Navigate relies solely on SIS data (demographics, academic history, academic performance history, etc.). Tools using this data generally can identify curricular choke points for individual classes and for sequences of classes. Staff involved in conducting institutional research, faculty and staff engaged in student success efforts such as SSI 2, and Academic Success Coordinators use the EAB Navigate tool for curricular analyses, advising campaigns, and interactions (scheduling) with students. This tool has access to SIS data, including a view of multiple courses taken by an individual student. The model used in EAB Navigate was recently re-oriented to predict semester-to-semester persistence (it was previously focused on six-year graduation rates). It provided an 87 percent success rate in predicting persistence (see Appendix N). Currently, academic advisors are being encouraged to prioritize student caseloads with an eye toward supporting students who appear to be in need of greater assistance.

**CANVAS LA DASHBOARDS**

These dashboards rely solely on in-course Canvas data (student performance data drawn from the Canvas grade book, interactions with discussion groups, assignment submissions (or the lack thereof), and login history, among other data sources). There are two dashboards in Canvas courses—the classic dashboard, and a proposed (and improved) dashboard that shows promise. Neither tool extends across multiple courses, however, and neither is generally available to advisors.
COMBINED SIS/CANVAS LA DASHBOARDS - LOUDSIGHT AND HOME-GROWN OR UNIZIN PARTNERSHIP OPPORTUNITIES

Unizin systems and third-party tools are beginning to draw together data from both SIS and Canvas to create richer insights into student learning. These technologies provide a window into student performance (data drawn from Canvas and related tools) that is informed by a student’s academic history and demographic background (data drawn, in CSU’s case, from Banner). In addition to descriptive analytics, this combination of data can facilitate the development of robust predictive analytics.

CSU piloted LoudSight, an analytics solution initially developed by the startup company LoudCloud and subsequently acquired by Barnes and Noble Education (BNED). The pilot studies were conducted over the past three years, initially under a direct contract with LoudCloud and then through a contract between Unizin and BNED. The Unizin contract was terminated in June 2019. While LoudSight provided an extensive and comparatively sophisticated toolset, the product was eventually deemed not ready for large scale use. However, the LoudSight pilots prompted greater faculty and advisor interest in the combination of SIS and LMS data in a single, easily-accessed location. Prior to cancellation of the contract, three pilots of LoudSight were completed at CSU, and all of the instructors involved expressed a desire to retain access to the kind of analytics provided. CSU is currently exploring alternate tools as well as the potential of a local programming/development project or a collaboration between Unizin institutions for a shared solution. The knowledge gained from the LoudSight pilots will inform selection of future dashboards and nudging tools.

Moving forward, the plan is to use the Unizin Data Platform (UDP) to retrieve SIS, Canvas, and third-party in-course data (e.g., data from iClicker, Kaltura, and products that report on student reading, viewing, and use of adaptive course materials). This approach will relieve Information Systems (IS) from the burden of populating an LA tool with SIS data every term, expand the available dataset, and ease data transfer processes. (IS has exported SIS data into the UDP encompassing data mapping, coding, and automation.) The availability of in-course Canvas data and SIS data together in one system increases the effectiveness of our instructional and advising interventions. It also enables research on student learning, student success, and new LA tools and processes.

STUDENT TOOLS

Unizin members have created new ways to (1) engage students with their own data and (2) provide nudges around assignment submission. These universities have expressed their willingness to partner with others using the UDP. Unizin is exploring how best to share the code from these development projects. These new tools include:

- Indiana University’s Boost: https://boost.iu.edu/
- University of Michigan’s My Learning Analytics (MyLA): https://sites.google.com/umich.edu/my-learning-analytics-help/
- University of Iowa’s Elements of Success (EoS): https://teach.uiowa.edu/elements-success
Explorations have begun on a shared, collaborative programming/development effort among interested Unizin institutions.

**CURRICULAR ANALYTICS**

A new tool for curricular analytics (appropriately named “Curricular Analytics”) is being explored by CSU staff to identify inordinate and unneeded complexity in curricula [https://curricula.academicdashboards.org/](https://curricula.academicdashboards.org/). Two metrics are used to measure complexity: (1) structural complexity in the relationship between courses in the curriculum and (2) instructional complexity as measured by high numbers of D/F/W grades in courses. This tool complements and supplements some of the analyses in EAB Navigate. The outcomes of the pilot may lead to further analysis as to implementation.

5. Ethical Uses of Learning Analytics

Ethical issues associated with learning analytics include both reasons to use LA data and tools and reasons to use them with caution. Ethical arguments in favor of using LA data and tools rest largely on the principle of acting in the best interests of students and instructors as well as the larger Colorado and national communities. Simply put, failing to follow up on opportunities to enhance student learning and success works against the interests of our students; failing to provide information that would develop stronger teaching practices works against the interests of our instructors; and failing to provide the largest number of capable, well-educated students works against the interests of the communities CSU serves.

Arguments favoring a cautious approach to using LA data and tools rest largely on concerns about the maturity of the tools, the reductiveness of information they provide, and potential abuses of student privacy and faculty academic freedom. While recognizing the insights afforded by the use of LA tools, a number of scholars have called attention to the potential misuse of information LA tools produce. Sharon Slade and Paul Prinsloo (2013), for example, observed that predictions about the likelihood of successful course completion could lead instructors and advisors to discourage students from taking courses or pursuing programs of study in which they are likely (but by no means guaranteed) to fail. Their caution is particularly important given the difficulty faced by students—often first-generation college students and/or members of historically underrepresented groups—who might enter higher education courses with comparatively lower levels of academic preparation than students who are members of families that enjoy higher socio-economic status or families that include members with college degrees. Slade and Prinsloo also expressed concern that inappropriate conclusions might be drawn about the teaching effectiveness of faculty members, a concern that echoes arguments made by a number of scholars about the reductive nature of student evaluations of teaching (see, for example, the 2017 meta-analysis by Uttl, White, and Gonzalez). Other scholars have argued that LA tools are too immature to be used without a great deal of caution, citing privacy concerns (Jones & Salo, 2018; Pardo & Siemens, 2014), reservations about issues related to privacy and the potential commercialization of student data (Flavin, 2016; Rubel & Jones, 2016), and concerns about the reductivism inherent in any analysis of “big data” (Stephens, 2017).
The importance of these concerns for scholars involved with learning analytics are addressed in the editor’s introduction to a recent issue of the *Journal of Learning Analytics*.

Questions related to privacy and ethics in connection to learning analytics have been an ongoing concern since the early days of learning analytics. Examples of some of the major questions are related to the ownership and protection of personal data, data sharing and access, ethical use of data, and ethical implications of the use of learning analytics in education. It is well recognized that these issues lie at the very heart of the field and that great care must be taken in order to assure trust building with stakeholders that are involved in and affected by the use of learning analytics. (Gašević, Dawson, & Jovanović, 2016)

With these concerns in mind, numerous proposals have been made regarding ethical principles and practices related to both the analyses that LA tools produce and access to the data on which they are based. In 2013, George Siemens suggested that we look not only at data ownership and retention but also at the issue of learner control over how their data should be used. One year later, Abelardo Pardo and Siemens (2014) proposed an ethical framework for learning analytics that focused on four aspects of privacy that had emerged in response to the growing collection of digital user data over the past two decades: “transparency, student control over the data, security, and accountability and assessment” (p. 448). More recently, Andrew Cormack (2016) has argued that we should draw on ethical frameworks used in medical research to separate “the processes of analysis (pattern-finding) and intervention (pattern-matching)” so that we can protect learners and teachers from “inadvertent harm during data analysis” (p. 91). Hendrik Drachsler and Wolfgang Greller (2016) proposed DELICATE, an eight-point checklist based on recent legal principles and the growing literature on ethical use of LA data that supports a “trusted implementation of learning analytics” (p. 89). And in a promising approach to preserving privacy while ensuring benefits to learners and teachers, Mehmet Emre Gursoy, Ali Inan, Mehmet Ercan Nergiz, and Yucel Saygin (2017) have developed and tested a framework for the development and enforcement of “privacy-preserving learning analytics (PPLA)” (p. 69).

Building on these efforts, a small but growing number of higher-education institutions (e.g., Charles Sturt University, 2015; Colorado State University, 2018; University of Michigan, 2018), professional organizations such as the Society for Learning Analytics Research (Gašević, 2018) and the Reinvention Collaborative (Jensen & Roof, 2017), and non-governmental organizations such as Jisc (Sclater, 2014; Sclater & Bailey, 2015) have developed frameworks to inform the ethical use of LA data and tools. Other institutions and organizations are currently adapting existing or developing new frameworks.

In our considerations of ethical issues related to learning analytics, the task force identified the following areas of concern.

**USING PREDICTIVE ANALYTICS**

For instructors, predictive analytics (particularly “zero-day analytics,” which are provided prior to the start of a course) have the potential to set up biases about student abilities and potential. Our recommendation is to pursue two paths. First, prior to start of the course, we would be wise to report predictive analytics to course instructors only for groups of students (see Recommendations 12 and 13), and then only after instructors have completed some initial training on how to use such information in
educationally effective ways. For example, instructors might be informed about the percentage of first-
generation students, majors/non-majors, and racially-minoritized students in their courses. They might
also be provided with the range of GPAs and other descriptive performance indicators. We suggest that
these data be accompanied by links to teaching strategies shown to be effective with the relevant
student groups. At some point in the semester, predictions based in part on these and other
demographic and academic factors might be applied to individual students. It could be useful to see
which students are performing above or below predictions, for instance, so that instructors can identify
students in need and get a sense of how the class as a whole is performing. This approach is in use at UC
Davis, where Carolyn Tomas, Vice Provost & Dean for Undergraduate Education, and Marco Molinaro,
Assistant Vice Provost for Educational Effectiveness, have been developing a model for providing and
using student data that requires faculty members who wish to obtain such data to complete relevant
trainings.

Second, recognizing that bias is often the product of lack of knowledge about the limitations of the
information provided to us, we recommend that professional development efforts regarding the use of
learning analytics include significant attention to the limitations of predictive analytics (see
Recommendation 13), the potential for implicit bias, and appropriate interpretations and use of the
data. With an understanding of those limitations, instructors will be better prepared to interpret and
use the information provided by these tools appropriately, both prior to and during a course.

In addition, it’s clear that we need to identify appropriate interventions—or at least gain a better
understanding of previously-identified productive behaviors that might be recommended to students
who exhibit particular patterns of behavior (or lack of behavior).

For advisors, it will be useful to consider how predictive analytics might be used to channel students into
paths other than those students desire to take. For example, by discouraging students from pursuing
programs of study in which they are likely (but by no means guaranteed) to fail, predictive analytics
tools could serve to restrict opportunities for students based on background and experience rather than
ability and potential.

With these considerations in mind, it might be the case that bias will be avoided through the careful
development of nudges and alerts that can be sent automatically rather than solely at the discretion of
an instructor or advisor (see Recommendation 4). Work in this area might eventually allow us to use
nudges and alerts in ways that are more equitable than current practices.

**ASSESSING TEACHING EFFECTIVENESS**

Administrators—in particular, department chairs/heads and college deans—might reasonably assume
that LA data and tools can provide insights into the performance of individual instructors and advisors.
Similarly, they might be tempted to use these tools to gain insights into instructors as a course is being
offered, with the intention of providing formative feedback while it can benefit student learning and
success. We recommend that information from these tools be used, if it is used at all, only in
conjunction with other information about teaching effectiveness. It should not serve as the sole or
primary basis for assessing teaching effectiveness (see Recommendation 9).
USING MULTI-MODAL DATA

Multi-modal data, such as location/time data revealed through connections to Wi-Fi routers or harvesting of social media behaviors and posts, can help us gain insights into student learning behaviors outside the classroom. This kind of data can help us learn, for example, which students are attending tutoring or study group sessions or visiting the library, and how much time they spend in these activities.

While we can collect this kind of data, we should avoid doing so. And in cases where collection is a normal part of the process of gaining access to a resource (such as logging into a Wi-Fi router), we should not include this data in our analyses. We believe strongly that student privacy should be respected and that the collection of this data, while potentially useful in developing and applying models of behaviors that lead to student learning and success, is both intrusive and unnecessary.

This does not mean, however, that we should not rely on data that is provided by students as they use university services, such as tutoring, sponsored study groups, and advising. Students are aware that their attendance is recorded at these and similar activities.

WORKING WITH LA TOOLS VENDORS

Vendors, including publishers and other learning companies such as EAB, will have access to significant amounts of data about our students. This raises numerous ethical and regulatory challenges for the university. These relationships should be a constant area of focus in our learning analytics efforts. We recommend that all faculty and staff who adopt LA tools should be aware of the requirements for vendors to complete the Colorado State University Digital Tool FERPA, Data Ownership, and Data Privacy Agreement (see Recommendation 9). Because most faculty may be unaware of the extensive—and substantive—implications of vendors’ uses of students’ data (and possibly personally identifiable information (PII)), we recommend that guidelines be developed to aid faculty and staff in making informed choices regarding the use of e-texts, adaptive courseware, and other vended digital learning platforms (see Recommendations 10 and 11). These guidelines should address issues including vendors’ potential use of students’ data to inform the development of new products, their use of students’ data for sale to third parties, and vendors’ potential use of faculty and members’ intellectual property (in particular, course materials posted on vendor sites).

6. Professional Development for Users of Learning Analytics Data and Tools

CSU should design, implement, and assess a professional development initiative that helps faculty and staff understand how to use LA data and tools wisely and appropriately (see Recommendation 13). This initiative should be carried out by TILT with input and involvement from Faculty Council, the Office of the Provost, the Office of the Vice President for Student Affairs, the Office of the Vice President for Research, the Office of the Vice President for Diversity, and the colleges. The initiative should focus on use of LA data and tools to plan, conduct, and assess courses; the use of insights from LA data and tools to support the development of equitable and inclusive instructional practices; the use of messaging tools to enhance student learning and success; and the limitations of the tools. The initiative should also

---

2 Some learning analytics researchers are also referring to audio and voice recordings as well as biometric data as “multi-modal.” These kinds of data, when used to assess learning, are often collected in lab settings.
focus on CSU policies for working with LA tools vendors. The initiative should be assessed for long-term effectiveness and impact, with an eye toward determining whether the professional development activities are effective in combating biases. The overall goal of the initiative should be to ensure that LA tools are used to encourage student success going forward in their courses, as opposed to identifying deficiencies that may be inferred from demographic data and academic history.

The professional development initiative should align with TILT’s recently developed Framework for Teaching Effectiveness (see Appendix H). The framework takes a comprehensive approach to helping instructors develop teaching effectiveness by organizing professional development opportunities into seven categories: instructional strategies, curriculum/curricular alignment, classroom climate, pedagogical content knowledge, student motivation, inclusive pedagogy, and feedback and assessment. Already, some LA tools and data offer insights that can help to increase effectiveness of curriculum/curricular alignment, instructional strategies, and pedagogical content knowledge (e.g., of which course content tends to pose greatest difficulties for students and how to best support students in mastering it).

Professional development offerings in these categories, and others informed now or later by LA data, should be designed and made available as part of the comprehensive set of TILT faculty professional development to be provided through the framework. Such offerings should draw substantively on relevant research, for instance by asking faculty and staff associated with C-ALT, CSU Online’s Research & Analytics, the Office of Institutional Research, Planning, & Effectiveness, and colleagues with relevant expertise to consult in their design. These offerings should include workshops, seminars and other development opportunities with substantial online components available on demand as part of a larger suite of planned hybrid faculty and staff development offerings. Further, to encourage faculty members and staff to participate in these offerings, TILT should include them in the categories now being defined as part of a Provost’s Academy project that will offer faculty Provost’s Office recognition for initial, continuing, and sustained engagement with professional development in teaching effectiveness.

7. Coordinating Learning Analytics Efforts at CSU

For the past five years, the Learning Analytics Steering Committee, which has drawn its membership from the Provost’s Office, Faculty Council, TILT, CSU Online, and C-ALT, has served as a coordinating body for a set of LA initiatives across campus. This work has involved regular meetings, consultation with involved units across CSU, coordination with Unizin efforts, coordination with the Canvas Coordinators, and a summer retreat on LA.

Models used at other universities include assigning responsibility for LA efforts to a vice provost or associate provost, assigning responsibility to a central computing or IS group, assigning committee to a new central group charged with LA efforts, and the grass-roots model that was used at CSU prior to the formation of the LA Steering Committee. The members of the task force considered these models. Given the relative fluidity of developments in LA, we recommend that we continue to use the steering committee structure for at least three more years, at which time the question of other models should once again be considered.
8. Recommendations

LEARNING ANALYTICS PURPOSES AND INFRASTRUCTURE

1. Learning analytics data and tools should be used to enhance student learning and success.

   CSU should use learning analytics data and tools primarily to understand students’ learning behaviors, identify opportunities to promote effective behaviors as a course is offered, and to inform subsequent development and improvement of curricula and instructional materials. This understanding should be used to promote effective learning behaviors. CSU should also use insights gained through improved understanding of students’ learning behaviors to inform the development of improved professional development initiatives.

2. The existing and emerging LA dashboards in Canvas should be promoted as a default LA environment for all Canvas courses, while more advanced tools should be promoted for use in selected instructional areas.

   Using the LA tools in Canvas offers an easily-implemented path to wider use of LA to enhance teaching and learning at the university. Doing so will require significant effort, however, in the area of faculty and staff professional development. We recommend that instructors receive increased training in the use of tools within the Canvas LMS. This professional development effort will require significant and extensive effort on the part of TILT, ACNS and the Library, the colleges, and the departments.

   To gain greater insights into learning behaviors and instructional innovations will require a more robust set of tools than those offered through Canvas. For the past three years, CSU has piloted the use of LoudSight. Now that the contract between Unizin and BNED has ended, LoudSight is no longer available. However, we have gained important insights into the use of LA tools such as those provided by LoudSight. These insights can be applied to our use of Canvas LA tools. They can also be applied to our use of new tools that are being developed at Unizin partner institutions and, in the coming years, can be deployed and used at CSU.

   The LoudSight tool has received strong interest from all faculty who have used it in our pilot projects. Advanced tools, in place of or used in conjunction with the Canvas LA tools, should be used to inform our understanding of the results of key instructional initiatives, such as interventions in lower-division courses with high rates of D/F/W grades. To be most effective, the use of these LA tools should be coordinated with related efforts, such as the integration of active learning and instructional approaches based in findings from cognitive psychology research on how people learn.

   We recommend that CSU pursue one or more of the following options: (1) the adoption of one or more LA tools developed at other Unizin institutions, (2) participation in a consortial LA tool development project involving other Unizin institutions, or (3) the development of a home-grown platform that draws on code created by other Unizin institutions that should be contributed back to Unizin. We anticipate that each of these options will require funding of at least one FTE programming position.
3. **CSU should continue to use EAB Navigate to support our student success efforts.**

EAB navigate is ensconced in the ASC environment, and is necessary as a tool for the enhanced success of students under their purview. The Student Success Initiatives, and in particular ASCs should explore whether they intend to use predictive analytical tool to identify an individual student’s success in a major, and whether they have a need for “zero-day predictive analytics” to identify an individual student’s success in a specific course. If so, an exploratory project on this topic should be defined, and a committee constituted and charged.

The Curricular Analytics tool to reduce major complexity should continue to be explored by academic departments in concert with IRP&E and the Provost’s staff.

4. **CSU should use LA data and tools to support and automate Early Performance Feedback efforts.**

The growth of the Early Performance Feedback (EPF) program can be supported through the use of an automated tool. Such a tool provides a scalable resource. Ideally, this tool would connect SIS and Canvas data with an automated messaging system and a dashboard visible to resident assistants and advisors. Based on a set of instructor-configured variables (e.g., current grade below 70 percent), alerts and notifications would automatically be available to students, advisors, and resident assistants.

Faculty participation in EPF has been linked to improved student success outcomes. With this in mind, we believe that integrating in-course data will help us expand use of the EPF program. Care must be taken to ensure that EPF use of LA data and tools will not unintentionally bring about negative responses from students and instructors, particularly in areas related to academic ability or the likelihood of success. Particular attention to the impact of such efforts on students of color, low-income students, and first-generation college students is advisable.

5. **CSU should use automated messages (nudges and alerts) sent through LA tools to encourage student learning and success.**

We recommend that automated messages should:

- Use positive framing that emphasizes the student’s potential for learning and academic achievement.
- Emphasize concrete actions the student can take to improve learning (e.g., use spaced or interleaved practice, self-explain key concepts, attend tutoring, study with peers, etc.).
- Include any available data on which behaviors have helped previous students with similar academic and demographic profiles succeed in the course (stated concisely in accessible language).
- Link to relevant resources (e.g., short videos on science of learning study approaches, Canvas listing of instructor office hours, tutoring schedule).
- Use student-friendly language, strictly avoiding negative impact on minoritized populations such as women and students of color in STEM fields.

These tools are both student-centric and course-centric, and they have the potential to enhance student learning in individual courses. We should deploy these tools in a way that involves
instructors, advisors, and/or resident assistants, with the goal of determining which of these groups might best respond to specific types of messages. We should also explore how students react and respond to nudges and alerts with the goal of improving the effectiveness and appropriateness of the messages. Our larger goal should be to create a culture in which all participants regularly use LA data to support learning and academic achievement. To accomplish this larger goal, we should:

- Be transparent about the nature and reasons for using LA data, e.g., to help the instructor support students in learning more effectively and/or to improve the course over time.
- Be transparent with students about what types of data are collected, where and how data are stored, who has access to data, and how the threat of a data breach is mitigated.
- Explain, in plain English, how LA algorithms are applied within our learning environment.

6. **CSU should continue to explore the efficacy and ethics of using zero-day predictive analytics in courses.**

Members of the task force expressed a range of opinions about the use of zero-day predictive analytics by course instructors, in many cases pointing to potentially negative effects of pre-judging the potential success of students and in others pointing to the benefits of understanding the needs of a class as a whole. A growing consensus emerged about the value of presenting zero-day predictions at a level that does not identify individual students. For example, information might be provided about the likelihood of success in the course among students who are first-generation, and that information might be accompanied by links to professional resources about teaching and learning processes relevant to such students. This might, however, introduce bias, since instructors can view students’ first-generation status in Aries. Another option might be to provide the distribution of de-identified predictive success scores for all students in the course.

Task force members had greater consensus about the value of such information for advisors. Paired with professional development about the potential use and misuse of such data, access to this information was found to be useful for advising purposes.

**ETHICAL AND INFORMED USE OF LEARNING ANALYTICS DATA AND TOOLS**

7. **CSU should recognize its ethical obligation to use LA data and tools to enhance student learning and success.**

CSU strives to act always in the best interests of students, faculty, staff, and the larger Colorado and national communities. Failing to follow up on opportunities to enhance student learning and success works against the interests of our students; failing to provide information that would develop stronger teaching practices works against the interests of our instructors; and failing to provide the largest number of capable, well-educated students works against the interests of the communities CSU serves. Failing to use information (appropriately and within the ethical guidelines discussed in section 5 of this document and in concert with Recommendation 8, below) provided by LA data and tools would be an ethical lapse. It would undermine our efforts to pursue our mission and limit our ability to support our students, faculty, staff, and larger constituencies.
8. **CSU should commit to using LA data and tools in ways that adhere to principles outlined in the *Ethical Principles of Learning Analytics* document.**

   CSU should identify and use LA tools and data thoughtfully, in the context of effective teaching approaches. CSU should understand and avoid the possible discrimination documented in many social-science analytics projects (Munoz, Smith, & Patil, 2016; Prinsloo & Slade, 2017; Williams, Brooks, & Shmargad, 2018). We should use LA data in ways that are transparent, formative, inclusive, and supportive of CSU’s teaching and learning mission. Following the recommendations in the *Ethical Principles of Learning Analytics* (Appendix A), CSU should as a matter of policy provide information about the information collected through learning analytics data and tools and the purposes for which that information is used (see Recommendation 20, below).

9. **CSU should avoid using LA data as the sole or primary source of evidence of teaching or advising effectiveness.**

   While it might be tempting to use LA data as indicators of teaching effectiveness (as, indeed, it has been all too tempting to use isolated questions for the ASCSU Student Course Survey for evaluation of teaching effectiveness), this practice should be avoided. Similarly, it might be tempting to use LA dashboards and reports to gain insights into the performance of instructors as a course is being offered, with the intention of providing formative feedback while it can benefit student learning and success. Because of the limitations of current LA tools, this should also be avoided. Instead, LA data should be used to prompt reflection on and improvement of instructional approaches and curricula, just as the revised course survey is designed to be used. This approach might be promoted by linking it to the new Provost’s Office requirement that departments articulate specific criteria for evaluating non-tenure-track faculty members teaching, with the explicitly stated expectation that these criteria will also guide evaluations of teaching effectiveness for tenure-track faculty. The Provost’s Office has encouraged departments to pursue reflective approaches that demonstrate the instructor’s use of data for teaching improvement. The use of LA data should be included in such demonstrations.

10. **CSU should require LA tools and data solutions providers to conform to CSU policies on use of LA data.**

    A growing number of vendors, including publishers and educational technology companies, are offering LA tools and data solutions. Because their products and services result in the collection of student data, vendors whose products are used at CSU should complete the Colorado State University Digital Tool FERPA, Data Ownership, and Data Privacy Agreement. This agreement was approved by the CSU Learning Management System (LMS) Steering Committee in Fall 2018 (see Appendix M).

    Conformance with the terms of the agreement should be carried out by the Central Canvas Coordinators and overseen by the Learning Analytics Steering Committee.
11. CSU should publicize to faculty and staff its guidelines for working with vendors who collect data on students and instructors.

A process is in place for reviewing agreements with vendors who collect data on students and instructors. However, many faculty and staff may be unaware of the extensive—and substantive—implications of vendors’ uses of students’ data (and possibly personally identifiable information). We recommend that these guidelines be subject to regular review and, as required, revision. We also recommend that professional development (see Recommendation 13, below) for faculty and staff include attention to issues such as vendors’ potential use of students’ performance data for sale to third parties and vendors’ potential use of faculty members’ intellectual property posted on the vendors’ course sites. Finally, we recommend that the guidelines be circulated regularly to faculty and staff (e.g., with Provost’s Office messaging about completing book orders) to ensure awareness of their existence and importance.

PEDAGOGY AND BEST PRACTICES

12. With predictive analytics, instructors and advisors should focus on delivering positive messages that will promote student success, such that even if the predictions are not completely accurate, any ensuing actions only improve student learning and performance.

Instructors and advisors who use predictive analytics should ensure that all messages related to predictions of success in a course focus on positive strategies for achieving success. Our goal should be to provide our students with options for success and constructive pathways for accomplishing their goals as learners. We should never discourage students from pursuing paths they wish to follow. Instead, we should provide the needed support to help them evaluate options and pursue success.

13. CSU should provide faculty and staff with guidance in understanding and using LA tools to promote student learning and success, particularly where sensitive Personal Identifying Information is exposed. Instructors using visible PII should be required to complete a course in ethical use and behaviors associated with the exposure of such information.

CSU should provide faculty and staff with guidance for understanding and using LA tools, as well as for collecting, interpreting, using, and sharing LA data. It is particularly important for instructors to understand that, even if they do not use LA data provided by online learning tools, these data are in many cases still collected and used by the vendors producing such tools.

This initiative should be carried out by TILT with input and involvement from Faculty Council, the Office of the Provost, the Office of the Vice President for Student Affairs, the Office of the Vice President for Research, the Office of the Vice President for Diversity, and the colleges. The initiative should focus on use of LA data and tools to plan, conduct, and assess courses; the use of insights from LA data and tools to support the development of equitable and inclusive instructional practices; the use of messaging tools to enhance student learning and success; and the limitations of the tools. The initiative should also focus on CSU policies for working with LA tools vendors.
The initiative should involve workshops, seminars and other development opportunities with substantial online components available on demand as part of a larger suite of planned hybrid faculty and staff development offerings. The initiative should be assessed for long-term effectiveness and impact, with an eye toward determining whether the professional development activities are effective in combating biases.

14. Effective use of LA tools is predicated on the availability of meaningful data. To provide that data, CSU faculty should be trained in appropriate use of the Learning Management System and in related learning tools.

Information provided through Canvas and other LA tools will depend strongly on how instructors use Canvas to support teaching and learning. As more features and tools within Canvas are used, more data will become available and more accurate conclusions can be drawn from student use of Canvas. Deepening use of Canvas features and tools can be accomplished, in part, through greater use of instructional design staff and resources.

15. Detailed predictive LA should be targeted toward areas where it will be most beneficial, especially in lower-division courses with high DFW grade rates.

A new course development initiative should be developed for lower-division courses with high DFW rates. The initiative should involve departmental and faculty engagement (typically course coordinators), required training for instructors, and support for deployment when all conditions have been met.

PLANNING AND REPORTING

16. CSU should continue the Learning Analytics Steering Committee for at least three more years.

We recommend using this structure for at least three more years, at which time the question of other models should once again be considered. See Section 6 for additional discussion of this recommendation.

17. CSU should develop a strategic plan to align learning analytics efforts with our mission and student success initiatives.

The goal of employing LA tools, data and analysis techniques is to better the student learning experience in order to increase persistence and eliminate attainment gaps. We recommend an approach to using learning analytics in our student success efforts that is data-informed but not data-driven. We also recommend an approach that draws on learning analytics efforts in departments and units while simultaneously developing capacity centrally. This includes providing support for professional development efforts (e.g., in TILT), research initiatives (e.g., through C-ALT), and capacity building (e.g., in IRPE, TILT, and CSU Online).

18. CSU should conduct a retreat to develop a LA implementation plan.

We recommend that CSU continue to organize retreats over the summer to identify how we should plan for deployments for the upcoming academic year and thereafter. These retreat should include
members of this task force and others who might be interested in LA data and tools, especially others who serve on the Learning Analytics Steering Committee. Principals from Unizin, APLU, WCET, EDUCUAS, and Unizin member institutions (in particular, Oregon State University, University of Michigan, and Indiana University) as well as other groups might be invited to inform discussions.

19. **CSU should document how we are using LA data and tools to enhance teaching, learning, and student success.**

CSU is committed to using LA data and tools in ways that enhance teaching and learning and promote success for all students, especially those from historically under-represented groups. That commitment should be documented on a public website that informs faculty, staff, students, community members, and stakeholders about how we are using LA to enhance learning, teaching, and student success.

This website should provide a concise, high-level overview of what is -- and is not -- known to date regarding the efficacy of using the major types of LA data to improve learning, teaching, and student success. It should also provide a bibliography of relevant research, links to reports from C-ALT, links to related material from Faculty Council (including the *Ethical Principles of Learning Analytics* document), and links to related work outlining best practices for using LA data and tools in the pursuit of improved teaching, learning, and student success.

Syllabi in courses that make use of LA data and tools should provide a link to this site. Similarly, all professional development offerings (e.g., workshops, short courses, seminars, etc.) for faculty and staff should share the link to this site and include a brief overview of relevant research and how it informs the material presented and approaches taught.

**RESEARCH**

20. **CSU’s use of LA data and tools should be informed by research.**

There is a substantial research basis for using LA data to improve learning, e.g., in Carnegie Mellon University’s [Open Learning Initiative](http://www.openlearninginitiative.com) adaptive courses. However, much is yet to be learned. Vendors’ development of platforms using LA data has outpaced research showing how particular uses of LA affect students’ learning-related behaviors and academic achievement or instructors’ effectiveness. Many platforms are not research-based. Many uses of students’ LA data for commercial and research purposes pursue goals other than improved learning.

At CSU, we should explore how LA data and tools can be used to enhance and deepen student learning. This research might be carried out by existing units, such as IRPE and TILT, as well as by faculty and students through C-ALT. Of particular interest are studies of alternative algorithms for assessing and predicting student learning and success and their utility in instruction and advising.

C-ALT should continue to conduct research projects along the lines it does now. We view these as having longer-term, strategic value, whereas the recommendations from this subcommittee primarily affect our operational environment.
Task Force Members

- Ryan Barone, Assistant Vice President for Student Success
- Melody Brake, ACNS
- Pat Burns, Vice President for Information Technology and Dean of Libraries
- Sean Burns, CSU Online
- Dan Bush, Vice Provost for Faculty Affairs
- Anne Cleary, Psychology
- Stephanie Clemons, Design and Merchandising
- Kimberley Corwin, CSU Online
- Gaye Digregorio, Collaborative for Student Achievement
- Bob Engmark, Director, Information Systems
- James Folkestad, Professor of Education; Director, Center for the Analytics of Learning and Teaching
- Gwen Gorzelsky, Executive Director, The Institute for Learning and Teaching
- Matt Hickey, Health and Exercise Science; Chair, Faculty Council Committee on Teaching and Learning
- Dave Johnson, Director of Research and Analytics, CSU Online
- Laura Jensen, Vice Provost for Planning and Effectiveness
- Stan Kruse, The Institute for Learning and Teaching
- Kelly Long, Vice Provost for Undergraduate Affairs
- Diane Miller, Computer Information Systems
- Kevin Nolan, ACNS
- Heather Novak, Institutional Research, Planning and Effectiveness
- Mike Palmquist, Associate Provost for Instructional Innovation (Task Force Chair)
- Amy Smith, Associate Provost, CSU Online
- Erica Suchman, Microbiology, Immunology and Pathology
- Becky Villalpando, Registrar’s Office

Report Submitted: July 31, 2019

References


Appendices

A. Ethical Principles of Learning Analytics at Colorado State University
B. Recommendations for Ethics Reviews of Learning Analytics Projects
C. Learning Analytics Tools for Advising and Predictive Analytics
D. Currently-Used Sources of Learning Analytics Data
E. Additional Areas that Could Provide Learning Analytics Data
F. Features to Consider as CSU Builds Its Learning Analytics Infrastructure
G. Fundamental Questions for Learning Analytics Deployments
H. Teaching Effectiveness Framework
I. Observations on Current Learning Analytics Tools
J. Learning Analytics Issues
K. Learning Analytics Task Force Guiding Questions
L. Colorado State University Digital Tool FERPA, Data Ownership, and Data Privacy Agreement
M. Student Success Predictive Model Report – Colorado State University
APPENDIX A. ETHICAL PRINCIPLES OF LEARNING ANALYTICS AT COLORADO STATE UNIVERSITY

April 3, 2019

To: Tim Gallagher, Chair, Faculty Council
From: Matt Hickey, Chair, Committee on Teaching and Learning
Subject: Report from the Task Force on the Ethical Principles of Learning Analytics

The Committee on Teaching and Learning submits the following report from the Task Force on the Ethical Principles of Learning Analytics.

**Background:** In the Fall 2017 term, CoTL charged a Task Force with developing recommendations regarding the institutional approach to the application of learning analytics. The charge arose from discussions with multiple parties on campus, including the VPIT office, ACNS, the Registrar, the Research Integrity office, and individual faculty. Following receipt of the report, CoTL has sought stakeholder input; the draft has been reviewed by Dr. Pat Burns, staff in ACNS, staff in the registrar’s office, and was discussed at an IRB retreat in Fall 2018 that addresses data safety and data privacy issues.
Ethical Principles of Learning Analytics at Colorado State University

A report created by the CoTL Task Force on the Ethics of Learning Analytics

Task Force Members:

Tim Amidon (English)
Steve Benoit (Mathematics)
Ben Clegg (Cognitive Psychology)
Gaye Digregorio (Collaborative for Student Achievement)
James Folkestad (School of Education) - (Task-Force Chair)
Moti Gorin (Philosophy)
Gwen Gorzelsky (The Institute for Learning and Teaching)
Matthew Hickey (Health and Exercise Science)
Laura Jensen (Institutional Research, Planning, and Effectiveness)
Dave Johnson (CSU Online)
Mary Ontiveros (VP Diversity)
Mike Palmquist (Associate Provost for Instructional Innovation)
Mary Pilgrim (Mathematics)
Marla Roll (Assistive Technology Resource Center)
Chris Seng (Registrar)
Bayler Shubert (Associated Students of CSU)
Stephanie Yassa (Associated Students of CSU)

Task Force meeting dates (Fall 2017): August 29th, September 12th, September 26th, October 10th, October, 24th, November 7th, November 21st, December 5th

Introduction

Data science, and the specific application of data science in the context of the teaching and learning environment known as learning analytics (LA), involves the collection, measurement, analysis and reporting of data about learners, their behaviors, and their contexts (broadly defined). At its best, LA provides opportunities to employ evidence-based learning and teaching practices in pursuit of our educational mission. Inherently such opportunities are also coupled with significant ethical challenges.

In the essay “What is Data Ethics?”, Floridi & Taddeo (2016) observe that “the extensive use of increasingly more data—often personal, if not sensitive (big data)—and the growing reliance on algorithms to analyze them in order to shape choices and to make decisions, as well as the gradual reduction of human involvement or even oversight over many automatic processes, pose pressing issues of fairness, responsibility and respect of human rights”.

Learning Analytics Task Force Report, July 2019
The principles that are put forward in this report are based on the following foundational observations. First, at best, LA may inform and equip, but can never replace individual instructors or their interactions with students in the context of teaching and learning. Second, LA tools are not inherently good, or even neutral with respect to the learning environment. Like any educational tool, the unreflective application of LA can in fact harm students and the learning environment. Any “good” for educational aims depends on thoughtful and informed application of LA by responsible instructors. Third, at best, LA and the attendant algorithms may help inform the learning environment for students. No algorithm should be taken to wholly define an individual student, nor can LA be taken as “determining” any specific educational outcomes for an individual student.

The Purpose

The purpose of this report is to establish and clarify a list of ethical principles that will guide the implementation and use of Learning Analytics at CSU. We recommend that CoTL develop a code of practice and guide the creation of faculty professional development initiatives based on these principles. In addition, in an effort to understand and practice Learning Analytics with the highest of standards that reinforce our commitment to the Colorado State University System mission and to its Principles of Community, we want to recognize that changes in institutional policy may be needed to reinforce the ethical principles in this report.

CoTL’s Task Force for the Ethics of Learning Analytics

During fall 2017, the Faculty Council Standing Committee on Teaching and Learning (CoTL) charged a task force to propose guiding principles for the design, development, and implementation of tools and technologies that employ LA on the CSU campus. The task force members were cautiously optimistic about the potential of LA to support teaching and learning. This cautionary tone grew stronger over time as the task force members deliberated on the ethical challenges presented by such work. This document is the result of that ongoing work and outlines principles that the task force considers essential for the ethical use of these advanced approaches on our campus.

The task force is convinced that decisions about how LA and related educational technologies are brought into learning environments are fundamentally decisions about CSU’s community and educational mission; LA is not merely a “technology choice.” Furthermore, these approaches are increasingly amorphous and are being employed across a spectrum that ranges from enterprise tools adopted campus-wide to the unique tools and techniques deployed in a single classroom. Moreover, LA and the attendant student data are not simply confined to the CSU environment; student data can move into vendor databases, where the subsequent uses of LA data, privacy protections, and ownership may not be clear. Deployment of LA at all levels can impact the community.

The Principles of Learning Analytics are currently under active development and review. Given this ongoing development, the committee is actively looking for critical review and feedback on this document. We recommend that CoTL continues to seek feedback from all stakeholders within our
community including but not limited to students, faculty, staff, and administrators. We encourage all stakeholders to provide feedback. When seeking feedback we suggest that you consider and share with stakeholders the following tenets that arose repeatedly during our task force meetings and that guided the development of the language of the principles.

1. These ethical principles are intended to be a foundational component of our institutional ethos.
2. These ethical principles are intended to be congruent with our Principles of Community.
3. These ethical principles should guide the selection of Learning-Analytics technologies that are used on our campus.
4. These ethical principles should guide the application of Learning Analytics methods (including but not limited to technologies, algorithms, tools, and interventions) used in our institutional educational endeavors.
5. Evidence-based research is central to understanding the impact of LA on our community. We will hold this tenet central to all LA based projects, applying methods that adhere to the rigors of open science methodologies. It is critical that these methods, projects, and tools be open to critical review and evaluation.
6. LA resources and research efforts should be used with special attention to enhancing educational attainment opportunities for those most vulnerable within our community. Consistent with our ongoing Student Success Initiatives, projects should be selected that, when successful, will improve the learning opportunities for vulnerable populations and the entire community of learners at CSU.

Related CSU Policies:

This code of ethics has been developed with reference to and in support of the following principles, policies, guidelines and rules at CSU.

- CSU’s Principles of Community
- CSU Policy: Accessibility of Electronic Information and Technologies
- CSU Policy: Inclusive Physical and Virtual Campus
- CSU Policy: Environmentally and Socially Responsible Procurement
- CSU Policy: Information Technology Governance
- CSU Policy: Human Subjects Research
- CSU Policy: Central Administrative Data Governance Policy
- CSU Policy: Information Collection and Personal Records Privacy
- CSU Policy: FERPA
- CSU Policy: Research Data
- CSU Policy: Information Collection and Personal Records Privacy Policy
- CSU Policy: Information Technology Security Policy
- CSU Policy: Red Flags Policy
- CSU Policy: Information Technology Governance Charter (ITEC)
- Colorado Open Records Act (CORA)
The CSU Principles of Community clearly articulate the shared values of our institution. The task force has developed the ethical principles of LA to frame the issues at stake and provide a framework on which future decision-making, policy construction, and practice can build. These ethical principles are intended to ensure that LA and related technologies/approaches are designed to serve our community mission of access, research, teaching, service and engagement.

**Ethical Principles of Learning Analytics**

The Committee on Teaching and Learning acknowledges that Learning Analytics raise a number of ethical and legal issues (including privacy rights). Furthermore, the body of literature makes frequent reference to the imperative that institutions articulate clear guidelines on ethical considerations surrounding such aspects as the rights and dignity of individuals, as well as openness about processes and practices (Pardo & Siemens, 2014; Siemens, 2013; Slade & Prinsloo, 2013). The literature is equally insistent on higher-education institutions ensuring that their legal obligations are being met in relation to personal privacy, data collection, and information protection (Kay, Korn & Oppenheim, 2012; Siemens, 2013).

The Ethical Principles of Learning Analytics are the foundational principles that define the University’s approach to the use of Learning Analytics within CSU’s teaching and learning environment. These are more than guiding principles; they are best thought of as the core ethical foundations of Learning Analytics at CSU. All Learning Analytics practices must be consistent with these most basic governing principles.

**Principle 1: Learning Analytics serve the teaching and learning mission of CSU.**

Learning Analytics, and all information technology in support of the institutional teaching mission, will serve to enhance the teacher and student interaction, placing an emphasis on enhancing individual student learning opportunities and student success.
Principle 2: Learning Analytics serve the aim of Inclusive Excellence in the Learning Environment

Learning Analytics are designed for equity and inclusive excellence in our educational mission. As educational leaders, we are responsible for being mindful of how Learning Analytics may reinforce the exclusion or marginalization of historically excluded groups and guard against such misuse.

Principle 3: Learning Analytics is accountable to academic and institutional integrity

As scholars, educators, and learners we are accountable for understanding the implications of collecting, distributing, analyzing, and making decisions based on Learning Analytics data. This includes the implications of making decisions based on algorithms or statistics that are not disclosed to or understood by the user(s).

Principle 4: Learning Analytics data will be collected and maintained to understand specific pedagogical questions.

Learning Analytics data is collected from learning and teaching systems, retained, and utilized for the purposes of enhancing learning and teaching. Holding true to this principle, LA data will be collected based on predetermined pedagogical reasons, used for those reasons alone, and deleted after that data has served that specific use.

Principle 5: Learning Analytics operates with transparency and accountability

As scholars and educators, we will be fully transparent with students about what types of data are collected, where and how it is stored, who has access to it, and how the threat of a data breach is mitigated. In addition, faculty and administrators are obligated to provide a method for dialog and discussion about any LA assessments. The use of LA algorithms that can not be clearly understood will be avoided.

Principle 6: Learning Analytics data use arises from respect for the individual

All faculty, staff, and students at CSU are valuable members of the CSU community. The design and application of all Learning Analytics methods recognizes the individual dignity, rights, and responsibilities of all students as learners, engaged with faculty in pursuit of educational
excellence. Given this, the primary use of Learning Analytics should be formative, helping all students to understand and pursue excellence in learning and all faculty to pursue excellence in teaching.
APPENDIX B. RECOMMENDATIONS FOR ETHICS REVIEWS OF LEARNING ANALYTICS PROJECTS

Note: This appendix outlines a proposal that might be considered in subsequent planning efforts related to learning analytics.

All faculty- or staff-initiated uses of LA data outside institutional endeavors should be reviewed for incorporation of appropriate ethics provisions. (Institutional LA endeavors, such as Early Performance Feedback, are beyond the purview of this proposed review process.) Where LA data and practices are part of proposed formal research and scholarly activities, the proposed activities must be submitted to the CSU Institutional Review Board for review before the research activities are initiated.

To guide individual faculty members, informal faculty groups, and others who wish to use LA data to improve instruction, we suggest a two-pronged approach. First, we recommend emphasizing education regarding the potential and risks LA data afford. Education should take the form of a communications plan that disseminates to faculty, departmental administrators, and college leadership a concise overview of the data privacy, data stewardship, and algorithmic discrimination. This overview should emphasize that all uses of LA data require significant care regarding data security/privacy, regardless of the purpose for which the data are being used. The overview should link to CSU websites with further information. We suggest that the overview accompany Provost’s Office communications regarding the evaluation of teaching effectiveness, as appropriate. Similarly, we suggest that a white paper on choosing digital products for courses accompany Provost’s Office communications regarding the Higher Education Opportunity Act (HEOA), which are circulated each semester to inform faculty about federal regulations governing course materials usage.

Second, we propose developing a University LA Ethics Committee that will use discussions of specific LA use cases to develop institutional norms, policies, and practices that align with the CSU commitments articulated above and to provide feedback to faculty members seeking guidance on using LA data to improve learning and academic achievement. This committee will include the Chair of the Learning Analytics Steering Committee, the Vice Provost for Planning and Effectiveness, the Vice President for Information Technology, the Assistant Vice President for Student Success, and representatives from C-ALT, CoTL, the Associated Students of CSU, and The Institute for Learning and Teaching. The committee will meet one to two times per semester to discuss use cases brought by faculty seeking feedback, using the Ethical Principles of Learning Analytics as a guide to consider implementation and use of LA. Some questions the committee might discuss could include the following:

- Does the LA case serve the teaching and learning mission of CSU?
- What security provisions will be put in place?
- Does the LA case serve the aim of inclusive excellence in the learning environment?
- What analysis will be done? How, and is that analysis appropriate for the intended purpose?
- Are the learning analytics data being collected and maintained to address specific pedagogical questions?
- What broader benefits justify analyses and other uses of data?
- Does the LA case operate with transparency and accountability to CSU students?
- Does the LA case recognize and protect the individual dignity, rights, and responsibilities of all students as learners, engaged with faculty in pursuit of educational excellence?
By producing two types of documents, the committee will develop a body of work that will be used to shape norms, policies, and practices. The first document type will summarize concisely the feedback given to faculty members. The second type will summarize concisely issues the committee believes should be addressed in institutional practice. Both document types will be submitted annually to the Learning Analytics Steering Committee, which will develop a process for using these case summaries, over the next three years, to draft institutional policy and practice guidelines.
APPENDIX C. LEARNING ANALYTICS TOOLS FOR ADVISING AND PREDICTIVE ANALYTICS

A variety of tools are or were used by the CSU community to enhance both advising and predictive analytics.

EAB-Navigate

Academic Success Coordinators and other advisors are encouraged to use the EAB’s Student Success Management System Navigate and specifically the analytics which predict a student’s likelihood to persist to the subsequent semester using 40+ factors. Initially, the factors are predominantly gathered from the student’s application. As each semester passes, those are replaced by performance and enrollment data. CSU has chosen to bucket students by low, medium, and high support priority, representing 70%, 25%, and 5% of all students in a caseload respectively. Overall, the model is accurate at a rate of 87% in predicting a student’s likelihood to persist, with the most accurate predictions appearing for the highest performing students. The analytics produced using Navigate are also used to inform targeted student success interventions.

Taking Stock

The Taking Stock Survey is administered to all new students in both fall and spring semesters. The survey results are used by Residence Hall Assistants to inform individual conversations with each of the students on their floor(s). Those conversations are intended to support students in the early transition to campus. In 2018, the data were moved into Institutional Research, Planning and Effectiveness for use in predictive modeling of student persistence. Combined with other student data, preliminary results do indicate a utility of the survey data.

National Survey of Student Experience

The National Survey of Student Experience is administered to first-year and senior-year students. The most recent administration will be completed in June 2019. While there are many question responses associated with persistence, the most associated indicator is actual completion of the survey.

iClickers

Many course sections use iClickers to both understand content knowledge and attendance.

LoudCloud

LoudCloud pilots used in-course Canvas data to inform targeted outreach and automated nudges. These dashboards garnered interest from faculty and advisors and were an encouraging complement to Navigate due to the Canvas integration and nudging capabilities. BNED has chosen to no longer support LoudSight, and Unizin terminated the contract in June 2019.
APPENDIX D. CURRENTLY-USED SOURCES OF LEARNING ANALYTICS DATA

In addition to tools/surveys used in advising and predictive analytics, there are also a variety of student behaviors and milestones that are continually monitored because of their association with persistence.

Course Taking

CSU leverages the advising community to encourage students (barring extreme circumstance) to complete foundational math and composition courses as well as 30 credits in their first year. This campaign is based on analysis which found that students who complete 30 or more credits, including math and composition, within their first academic year have 67% higher odds of retention and 76% higher odds of graduation [IRPE, 2019]. Therefore, post-orientation, advisors outreach to students registered for less than 15 credits. Related, the Fall 2018 FTFT cohort was exposed to the “CSU Momentum Year-1” campaign encouraging students to:

1. Successfully complete 30 credits.
2. Successfully complete foundational math and composition.
3. Successfully complete 9 credits within their interest area.
4. Develop a productive academic mindset.
5. Engage in professional guidance such as a meaningful 1-1 conversation with faculty or staff member about big life questions.

Each of these five components has a CSU-specific or national research foundation supporting the activity. CSU will evaluate the impact of the activity in Fall 2019.

Early Performance Feedback

The Early Performance Feedback program has noted that if a student is identified in the first few weeks of a course by his/her faculty as not meeting executions yet, the odds of successful course completion are 91% lower compared to a student who is meeting expectations after controlling for course type. There is a 39 PP difference in the predicted rate of course success based on EPF indicator after controlling for the type of course. If a student is “meeting expectations” at week 4 they have an observed course success rate of 90.4% compared to a 48.3% success rate among students that are “not meeting expectations yet.” These data are shared with advisors and other staff on campus to conduct targeted outreach for support, and the data also feed into the U-Turn student success event.

Application

Data from a student’s application is used heavily in the first semester for predictive analytics as described in the EAB models. Remedial status and income are strongly associated with student success as measured by GPA and persistence.

Person Data Codes

Data codes to indicate student involvement in high impact practices are being finalized and already used to assess the association between completing these practices and persistence to graduation as well as GPA and salaries after graduation.
Card Swipe

Card swipe data is currently used to understand who attends tutoring and if attendance is associated with course success. This same type of analysis could also be integrated into predictive analytics.
APPENDIX E. ADDITIONAL AREAS THAT COULD PROVIDE LEARNING ANALYTICS DATA

Many opportunities for future campaigns and strategies exist. Ideas include:

- Math outreach for students who have not taken the MPE or are not in algebra and are STEM majors.
- Potential major declaration mismatch outreach not to dissuade but open a conversation about exploration, career vs. job, and purpose.
- Default semester or first year schedules for students admitted with recommended support to include math, composition, and other foundational courses.
- Analysis of student contact with ASCs/C4E scholar contacts using predictive modeling to determine ideal number and length of 1-1 meetings by semester with sub-group analysis.
- High Impact Practice/Experience participation, when, and with sub-group analysis.
- Scaling the use of in-course predictive/learning analytics via LoudCloud.
- LMS log in and page view analysis.
- Expand grade monitoring via Navigate, currently used with athletes, C4E, and Key.
- TILT and faculty intervention study, including course redesign, first-four weeks faculty initiative, and intergroup relations faculty training.
- Use of mobile network data and/or one-card chip data to track student patterns and behavior for predictive modeling (library visits, library visits, etc.).
APPENDIX F. FEATURES TO CONSIDER AS CSU BUILDS ITS LEARNING ANALYTICS INFRASTRUCTURE

This is a planning document that lays out tools, features, and capabilities that might be included in future learning analytics tools. While many of these features exist in our current tools, numerous gaps exist.

<table>
<thead>
<tr>
<th></th>
<th>Instructors</th>
<th>Advisors</th>
<th>Students</th>
<th>Researchers</th>
<th>Admin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dashboards</td>
<td>• Section</td>
<td>• Per Student</td>
<td>• By Class</td>
<td>• By Class</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>• Selected Sections</td>
<td>• All Advisees</td>
<td>• Across All Classes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Advisee Groups</td>
<td>• Historical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reports</td>
<td>• Section</td>
<td>• All Advisees</td>
<td>NA</td>
<td>• Section</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Selected Sections</td>
<td>• Advisee Groups</td>
<td>• Selected Sections</td>
<td>• Selected Sections</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• By Class</td>
<td>• All Sections</td>
<td>• All Sections</td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td>• Graphs of behavior</td>
<td>• Graphs of behavior</td>
<td>• Graphs of behavior</td>
<td>• Graphs of behavior</td>
<td></td>
</tr>
<tr>
<td>Tools</td>
<td>• Messages sent and opened</td>
<td>• Messages sent and opened</td>
<td>• Progress toward defined goals</td>
<td>• Messages sent and opened</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Actions taken in response to messages</td>
<td>• Progress toward defined goals (e.g., a grade)</td>
<td>• Actions taken in response to messages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>• Messaging</td>
<td>• Messaging</td>
<td>• Messaging</td>
<td>• Messaging</td>
<td></td>
</tr>
<tr>
<td>Tools</td>
<td>• Nudges</td>
<td>• Scheduling</td>
<td>• Scheduling</td>
<td>• Scheduling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scheduling (meetings, and ted into Outlook &amp; Google Calendar)</td>
<td>(meetings, and ted into Outlook &amp; Google Calendar)</td>
<td>(meetings, and ted into Outlook &amp; Google Calendar)</td>
<td>(meetings, and ted into Outlook &amp; Google Calendar)</td>
<td></td>
</tr>
<tr>
<td>Query Tools</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>How best to describe these?</td>
<td>NA</td>
</tr>
<tr>
<td>Database</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>SQL, MySQL, R, others?</td>
<td>NA</td>
</tr>
<tr>
<td>Design Tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**System Capabilities**: Dashboards, Reporting, Automated Nudges, Timed Messages, Email Reports, Output Reports in Excel and PDF among other formats

Can we break down the features of specific dashboards and reporting tools? For example, letting instructors know whether a student has completed an assignment and notifying students that an assignment deadline was missed.

Can we develop or adopt tools that allow us to carry out natural language processing? There has been a great deal of discussion of sentiment analysis (a psycho-linguistic tool/method), but there is much more that could be used to gain a more complete understanding of student attitudes and behaviors. We should study this and learn how it might help—and what we should avoid in light of the ethical concerns discussed elsewhere in this report.
APPENDIX G. FUNDAMENTAL QUESTIONS FOR LEARNING ANALYTICS DEPLOYMENTS

1. Predictions of retention, persistence and completion, at enrollment, later?
   a. Advising campaigns using Navigate analytics
   b. Other analytics (Laura)
2. Predictions of success in major, at enrollment, later?
3. Pre-course predictions of success in individual courses ("0-day analytics")
   a. LoudSight
   b. Separate deployment, using R?
4. In-course predictions of success in individual courses ("0-day analytics")
5. Do we want an automated tool for LA?
   a. To target resources toward where they will do the most good
   b. To prevent information which may be sensitive from human access, except in very special circumstances
6. Do we wish to encourage strong student behaviors using LA?
   a. Nudges (hand in your assignments on time)
7. How do we want to affect student performance using LA?
   a. Alerts (do “THIS” to perform XX better)
8. In-course dashboards, do we want dashboards for
   a. Faculty in-course (LoudSight, with SIS data)
   b. Advisors in-course (LoudSight, with SIS data)
   c. Students (in-course LoudSight, with SIS data)
   d. Canvas in-course (no SIS data)
   e. EAB outside of course (campaigns, appointments, advisor notes)
9. How do we integrate all of the above with Faculty—training, usage, communications, ethics?
APPENDIX H. TEACHING EFFECTIVENESS FRAMEWORK

From the TILT website at https://tilt.colostate.edu/proDev/tef/:

The TE Framework and accompanying tools offer an integrated, manageable approach by organizing teaching effectiveness into seven domains—aligned with questions on CSU's new course survey—which focuses on the learning environment (not the instructor).

The Teaching Effectiveness Design Guide includes the Teaching Effectiveness Framework which is comprised of seven essential, interrelated components of effective teaching and a collection of tools to improve student learning. The components are grounded in evidence-based teaching practices and the scholarship of teaching and learning. The guide provides administrative leadership with tools and language to measure teaching effectiveness, and allows faculty to choose their own reflective teaching practice journey by identifying and focusing on a strength and an area of growth.
APPENDIX I. OBSERVATIONS ON CURRENT LEARNING ANALYTICS TOOLS

This section contains the questions posed to and answered by the LA infrastructure subcommittee. It is included because of its potential use in addressing more targeted questions about LA tools and processes.

What are the implications of using the Canvas dashboard (e.g., no SIS data)? Is it sufficient for general instructors?

The new Canvas LA dashboard is dramatically improved from the previous version. It is our opinion that it is indeed sufficient for general instructors, as the deployment of a new tool to replace BNED LoudSight will be targeted, and not extended comprehensively on campus. We believe we should endorse the Canvas LA dashboard for general use and hold regular training/demonstrations of it (but infrequently, as the user interface is already so user friendly—it will be more important to convene faculty groups to discuss best pedagogical practices in its use).

What are the implications of using advanced dashboards for students, faculty, and staff/advisors?

We believe that the replacement for LoudSight (whether provided by Unizin, developed in collaboration with other Unizin institutions, or developed at CSU) should be deployed in a targeted fashion on campus, primarily focusing on large enrollment, lower division courses, primarily in courses with high D, F, and W grades. In order for the new system to be effective, the data needs to exist in Canvas, and specific SIS data also needs to be fed into the LA platform. It took significant time to feed SIS data directly to BNED for inclusion in LoudSight, and it took the BNED team some time to ingest it and formulate their predictive models. It is our expectation that as the Unizin Data Platform (UDP) matures, it will encompass both SIS and Canvas in-course data. With that in mind, we expect the new LA platform will have a more efficient data ingestion process.

Where are we with nudges and with alerts/interventions? Where do we want to be at in 18 months or farther out?

We recently completed our third pilot with LoudSight. Our first pilot was small, and successful. Our second pilot was unsuccessful due to platform problems and instabilities, and yielded no outcomes. Our third pilot scaled up to be quite large (almost 10,000 student course section enrollments), and was focused mainly on the stability of the platform. We expect to receive some anecdotal information from the third pilot on the effectiveness of nudges, alerts, and the use of the platform overall. However, we need to define a formal approach, project, communication plan, and evaluation/assessment plan for these factors as we scale up the use of a new LA platform for lower division courses. We should also integrate other aspects in as we progress, including the science of learning, high impact practices, effective course redesign, etc. That is, learning analytics must not be pursued in isolation, but integrated with other learning initiatives and activities.
What kind of LA data can be provided to support Early Performance Feedback? How could it be provided? At what cost (staff time, expenditures)?

We need to assess whether the triggers that exist in LS are sufficient to trigger EPF effectively, and have engaged very preliminarily with the Collaborative for Student Achievement on this topic.

What is the best direction to pursue with Canvas and the new LA platform (homegrown or otherwise) dashboards? What are the current limitations? What is likely to be coming down the pike? Where should we put our efforts with LA Tools?

We believe the optimal approach will be a bifurcated plan: 1) using the new platform in targeted lower division, high D, F, and W courses, requiring significant support of the instructors and working with the Collaborative for Student Achievement to bundle EPF in, and 2) recommending the Canvas LA dashboard for all other course sections, that requires minimal support. Either approach will require a close working interaction with TILT to engage faculty/instructors.
APPENDIX J. LEARNING ANALYTICS ISSUES

Issues that require attention as we progress with our development of learning analytics tools and practices include the following.

1. We need to run up gently to using the new LA platform to automate the Early Performance Feedback (EPF), and ensure that its triggers are effective as “defaults,” even though the “defaults” will be editable/changeable by faculty, both individually and holistically (e.g., by uploading a CSV file to Canvas or into the new LA platform). A critical element is how EPF will exist as an automated product—it could be self-contained entirely within the new LA platform, or it could be extracted from the new LA platform and loaded into Canvas.

2. We then need to normalize the class sections in Canvas using EPF to utilize the new LA platform (or maybe even Canvas) to automate the EPF function. Careful communication and interactions with the faculty will be needed. We may also need to design the new LA platform to generate the default EPF triggers to be more consistent with our needs. It will be important to work with the Collaborative for Student Achievement and TILT in this regard.

3. Currently, due to limitations at Instructure (the company owning and providing Canvas), comprehensive Canvas data are available only 36 hours or more after real time. Unizin is working with Instructure to get this down to on the order of four (4) hours, which would be timely for feedback and acceptable to us. A hybrid approach has been proposed that uses the Canvas live events feed and shows promise here.

4. We need to work with our colleagues in Unizin to enhance the predictive analytics models (zero-day predictive analytics, and in-course predictive analytics), including enhancing data used in the models, and possibly enhancing the algorithms used in the models. We also need to determine if the trigger for EPF is adjustable for the individual course section by the faculty member, and if not work to make it so. Moreover, we need to ensure that we populate the new LA platform earlier than was the case with LoudSight, in time to get zero-day analytics from LS, should we decide to use them. Currently, the models take a couple weeks into the semester to develop, but this is much too far into the class to be useful as zero-day analytics.

5. We need to incorporate in the predictive analytics models the additional data being captured by Unizin as it evolves. Currently, this includes: what students read via the Unizin Engage eReader platform, information from the TurnitIn anti-plagiarism platform, and information Unizin collects from Personal Response Systems (iClicker cloud version for CSU).

6. We are certain that we will need to work diligently with each and every faculty member to institutionalize automated LA, and cover the following topics: ethical use of LA data, LA data definitions, meanings and interpretations, ensuring the right data elements are included in Canvas in the “right” way (nuanced meanings), and in the new LA platform dashboards—definitions and use.

7. We need to buttress our data feed to the UPD to include more and better data from us, specifically item such as: time spent, frequency, and type of tutoring attended (topics) tied to students and course sections, participation in student study groups (time spent, frequency, and topics covered by the group), engagement in high impact practices, etc.

8. Using LA to improve learning—how do we define and collect key evidentiary data that demonstrates the effectiveness of using an LA platform, initially nudges and alerts and other functions and features as they evolve? Social norming theory, mostly from public health, notes that population-specific nudges with local data can be powerful to improve behavior. This is written about in other disciplines, less so in education, and this may be an opportunity for us. We have some experts on this topic on campus in psychology and other departments who should be consulted as this moves forward.
APPENDIX K. LEARNING ANALYTICS TASK FORCE GUIDING QUESTIONS

Learning Analytics Ecosystem at CSU: Next Steps

January 31, 2019
Pat Burns and Mike Palmquist

Goals:

1. Explore ways in which learning analytics (LA) data can be used to inform our analysis of institutional effectiveness, teaching effectiveness, advising effectiveness, student learning, student success, and student retention.
2. Identify and recommend best options for collecting and analyzing LA data.
3. Identify desired features of LA tools.
4. Identify and offer strategies for addressing issues related to ethical uses of LA data, including but not limited to student and faculty privacy.
5. Identify and recommend best options for increasing faculty, staff, and student understanding of best practices related to LA tools and data.
6. Recommend best options/practices for institutional organization of LA resources, tools, and staffing.

Working Group: Institutional Effectiveness

Members: Laura Jensen (convener), Ryan Barone, Stan Kruse, Heather Novak.

Key Topics and Questions:

1. Summary of current work: computing a posteriori student success metrics by various attributes: retention, persistence, and success
2. Summary of current and planned work: e.g., running EAB Navigate curricular studies, such as completion rates for students taking Math and Comp classes in the first year, etc.
3. Goals, Plans: What needs to change/continue to support measuring the effectiveness of implementing LA and advising strategies at an institutional level? What resources would be required (or what changes in current practices would be needed)?
   a. Do nudges and alerts in LoudSight improve student success? Other?
   b. Does the use of EAB Navigate correlate with enhanced student success?

Deliverables:
Within 6 Months (what can we report on now)
Within 18 Months

Working Group: Advising (EAB Navigate)

Members: Ryan Barone will lead this group. Members: Laura Jensen, Gaye Digregorio and Becky Villalpando
Key Topics and Questions:

1. Advising: Description/assessment of cases, alerts, and campaigns given current capabilities
2. Student registration, getting essential courses done in Freshman year
3. Other areas that could be explored:
   a. Early performance feedback
   b. Guided onboarding/default schedules
   c. Program selection: predictions of individual student success in major
   d. Degree planning
   e. Student admitted with recommended support priority
4. Goals, Plans: What are we trying to accomplish by providing LA, PA and EAB Navigate data to advisors? What kind of resources will we need to use these data effectively (e.g., early performance feedback)? How might this tie into other areas discussed in this document (LA Tools, ethical issues)?
5. How are LA and PA tools and strategies improving/deepening student learning? How do we know?
6. Implementation / Operation: Who manages these systems? Any changes in existing practices? Are existing practices sufficient?

Deliverables:
Within 6 Months (what can we report on now)
Within 18 Months

Working Group: Learning Analytics Tools (Unizin UDP, Canvas, LoudSight)

Members: Amy Smith and Pat Burns will lead the group.

Key Topics and Questions:

1. What are the implications of using the Canvas dashboard (e.g., no SIS data)? Is it sufficient for general instructors?
2. What are the implications of using the LoudSight dashboards for students, faculty, and staff/advisors?
3. Where are we at with nudges? Where do we want to be at in 18 months or farther out?
4. Where are we at with alerts/interventions? Where do we want to be at in 18 months or farther out?
5. What kind of LA data can be provided to support Early Performance Feedback? How could it be provided? At what cost (staff time, expenditures)?
6. Goals/Plans: What is the best direction to pursue with Canvas and LoudSight (or homegrown or other vendor) dashboards? What are the current limitations? What is likely to be coming down the pike? Where should we put our efforts with LA Tools?

Deliverables:
Within 6 Months (what can we report on now)
Within 18 Months
Working Group: Faculty and Staff Professional Development

Members: Gwen Gorzelsky (convener), James Folkestad, Matt Hickey, and Dan Bush

Key Topics and Questions:

1. Ethical use of LA data
2. Using LA to improve learning
3. Target lower division, challenging classes
4. Goals/Plans: What kind of program could be developed to support faculty, advisor, and student use of LA tools/dashboards? Can it be supported within current TILT funding?

Deliverables:

Within 6 Months (what can we report on now)
Within 18 Months

Working Group: Future: Tools, Functions, and Ethical Implications

Members: Mike Palmquist will lead the group. Other members include Stephanie Clemons, James Folkestad, Matt Hickey, Diane Miller, Erica Suchman (others are welcome).

Key Topics and Questions:

1. Desired functions, tools, reports for
   a. Faculty
   b. Advisors
   c. Students
   d. Researchers
   e. Admins
2. Ethical implications / guidelines
3. Goals/Plans: What are the optimum functions, tools, and reports (in a perfect world, in a more pragmatic sense)? What policies should we consider implementing?

Deliverables:

Within 6 Months (what can we report on now)
Within 18 Months
APPENDIX L. COLORADO STATE UNIVERSITY DIGITAL TOOL FERPA, DATA OWNERSHIP, AND DATA PRIVACY AGREEMENT

THIS DIGITAL TOOL FERPA, DATA OWNERSHIP, AND DATA PRIVACY AGREEMENT (“Agreement”) IS HEREBY INCORPORATED AND SUPERCEDES ALL OTHER AGREEMENTS BY AND BETWEEN THE SUPPLIER AND THE BOARD OF GOVERNORS OF THE COLORADO STATE UNIVERSITY SYSTEM (“UNIVERSITY” OR “University”), AS OF THE DATE BOTH PARTIES SIGNED THE AGREEMENT.

Background:
Supplier provides certain services and/or licenses certain applications to the University, often under separate contract(s), end user license agreement, privacy policy, etc. pursuant to which Supplier may have access to “education records” from the University, as that term is defined in the Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. 1232g, et seq. and the regulations promulgated thereunder (“Education Records”). To the extent there is a conflict between the terms of this Agreement and any other agreement or terms (including any terms of service or other click through terms necessary to access any Supplier services) the terms of this acknowledgement shall prevail.

Definitions:

University – Colorado State University (CSU).
Supplier – digital tool supplier.
GDPR – European Union’s General Data Protection Regulation.
Learning Management System (LMS) – the University uses the Canvas LMS.
Learning Tool Interoperability (LTI) – IMS Global Learning Consortium standard for integrating a digital tool with an LMS. Supplier LTI’s should be IMS certified.
Application Programming Interface (API): Tool data access done on behalf of a user in the LMS.
Canvas Developer Key: Developer keys can be used to create custom integrations with Canvas and allow third-party apps to use Canvas authentication. It is rare for CSU to issue a developer key.

“University Data” means all records and information created, received, maintained, or transmitted by the University which is accessed, created, used, stored, copied, or distributed by Supplier, in connection with the Work under the Contract.

“Data Breach” means, for the purposes of this Contract, any adverse event where there is harm to University Data, individuals, host(s), or network(s). This includes, but not by way of exclusion, events indicating that University Data may have been accessed, disclosed, or acquired without proper authorization and contrary to the terms of this agreement or the Contract.
Supplier and CSU hereby agree as follows:

1. **School Official:** To the extent that Supplier has access to University Data (Education Records) from the University, Supplier is deemed a “school official,” as that term is defined under FERPA. Supplier agrees that it shall not use Education Records for any purpose other than in the performance of the services under contract. Except as required by law, Supplier shall not sell, disclose, distribute nor otherwise share Education Records with any third party unless permitted by the terms of the contract or to subcontractors who have agreed to maintain the confidentiality of the Education Records to the same extent required of Supplier hereunder. In the event any person(s) seek to access Education Records, whether in accordance with FERPA or other federal or relevant state law or regulations, the Supplier will immediately inform University of such request in writing (email is acceptable).

2. **Data Ownership:** Supplier acknowledges that, as between the University and Supplier, all Education Records are and shall remain the property of the University; this includes any and all information or data that are collected, processed, viewed, stored, or transmitted either individually or in aggregate from CSU faculty, staff, and students.

3. **Data Use:** Supplier acknowledges that it shall use Education Records only as necessary to provide the applicable services or as otherwise permitted in writing by the University.

4. **Data Privacy:** Supplier shall not share any information obtained from University Users under this agreement with third parties, without obtaining explicit permission from the University or from the End User in question.

5. **Data Security:** Supplier shall maintain an approved and documented information security program to protect and safeguard Education Records, which shall include administrative, technical and physical safeguards that utilize commercially available industry best practices and comply with the requirements of FERPA.

6. **Data Access:** Where feasible, University asks that Supplier data adhere to the IMS Caliper Analytics standards to make it easier to incorporate into University’s data analytics process.

7. **Breach Notification:** If Supplier believes that any Education Records have been subject to unauthorized access, Supplier will promptly (and in any event within 48 hours) notify the University by sending e-mail to canvashelp@colostate.edu, detailing the nature, range, and possible scope of the unauthorized access. Any breach may be grounds for immediate termination of the Agreement by the University. Supplier shall work with the University, at the University’s request, jointly to isolate, identify and scope the nature and range of the breach, and discover the extent of the data exposure, data transmission, date receipt, etc.

8. **Data Retention and Disposal:** Supplier shall, on all of its systems, retain no (i.e. dispose of securely) information collected from University Users within two years past the end of the time it was used in a course, section or otherwise delivered service. Upon one year (360 days) after expiration of the Contract, Supplier shall consult with University to identify any and all information to be returned to the University, and subsequently delete all such information within one year past the expiration date of the contract/agreement.
9. **Termination:** Notification via email if otherwise verified or by FAX by University or Supplier of discontinuance of use or provision of this product shall be necessary and sufficient to effect termination of this agreement.

10. **Indemnification:** In accordance with State and University fiscal rules, the University is not allowed to indemnify and/or hold harmless any Supplier whatsoever for whatever purpose.

11. **Choice of Venue, Choice of Law:** In accordance with State and University fiscal rules, the University is not allowed to agree to a venue or choice of law, except for Colorado.

12. **Arbitration:** In accordance with State and University fiscal rules, the University is not allowed to agree to arbitration to settle any disputes.

13. **EU GDPR:** Supplier agrees to comply with the provisions of the GDPR at the University’s request, if practicable, including but not limited: collecting the minimum amount of information necessary to perform its duties, limiting retention of personal data as specified therein, deleting information specific to individuals upon request – except as necessary to perform its job function, and responding to requests from the instructor for the status of individuals’ data. Supplier shall assume responsibility for interacting with any and all individual inquiries to Supplier constituting a request under the GDPR.

14. **Disputes:** Any dispute arising from the services covered hereunder may be elevated for resolution to the Office of General Counsel, CSU, whose decision shall be final.

**Digital Tool Integration with Canvas**

Upon formal review, the University may choose to integrate a Supplier’s digital tool with its LMS, Canvas.

- **LTI:** The tool must be IMS LTI Compliant at version 1.0 or higher and working towards version 1.3 / LTI Advantage compliance.
- **Instructure Partner:** Though not required, CSU prefers to work with Suppliers who partner with Instructure to insure their product adheres to standards and interoperability with Canvas.
- **API Access:** If required, supplier will list what portion of the LMS is and which data are to be accessed via which specific API access points.
- **Grade Sync** – If the digital tool is authorized by CSU to push student assignment and grade information into Canvas, grade sync integration must meet CSU’s IT Security Policy (http://policylibrary.colostate.edu/policy.aspx?id=492) and Acceptable Use Policy (http://policylibrary.colostate.edu/policy.aspx?id=704).

**Developer Key** – Where applicable, in rare instances, the University will authorize and issue a Canvas Developer Key to facilitate deeper product integration between the Supplier and Canvas. A developer key authorizes the tool to make API requests on behalf of a Canvas user. If required for tool integration, Supplier must complete and return the CSU Canvas Developer Request Form for a key request to be evaluated.
If approved, supplier agrees:

- To use the developer key only for the expressed purpose identified on the CSU Canvas Developer Key Request Form.
- Identify any specific API endpoints that are needed.
- To diligently and securely store the key and strictly limit access to the key’s configuration information only to those with an immediate “need to know.”

If use of the developer key results in unintended changes or damage to CSU’s Canvas, CSU reserves the right to do any or all of the following:

- Work with the CSU to analyze, isolate, and repair any damage.
- Remove the supplier’s developer key from CSU’s Canvas.
- Remove the supplier’s LTI from CSU’s Canvas.
- Seek damages in proportion to the damage, downtime, lack of availability, etc. for all those involved in the breach.

In witness whereof, the legal extent and sufficiency of which is hereby acknowledged

SUPPLIER:_________________________________ FOR COLORADO STATE UNIVERSITY

By:_______________________________________ By:_______________________________________

Printed Name:___________________________ Printed Name:___________________________

Title:____________________________________ Title:____________________________________

Date of Signature:_______________________ Date of Signature:_______________________
APPENDIX M. STUDENT SUCCESS PREDICTIVE MODEL REPORT – COLORADO STATE UNIVERSITY

Executive summary

EAB has built a customized Student Success Predictive Model (SSPM) for your institution that predicts the persistence likelihood of your students. Your SSPM incorporates the latest breakthroughs in statistics and data science, placing your institution at the cutting edge of student-insight technology. It is a powerful tool for promoting your students because it gives you invaluable insight into their likelihood of academic success. This document provides an overview of the SSPM, describes how it was built and extensively customized and optimized for you, and details benchmarks of its predictive performance.

Performance Summary

The primary metric EAB uses to benchmark model performance is high-risk student identification rate. It is based on the most common use case for the model: that you are designing a campaign targeting high-risk students but only have the capacity to advise a limited subset of your total student population. In this case, your goal is to efficiently use your constrained resources to reach as many of your school’s actual high-risk students as possible.

The table below summarizes your SSPM’s performance and compares it to the following notional models:

- A fictitious, perfectly prescient model (Crystal Ball).
- A model based exclusively on students’ cumulative GPAs (GPA Model).
- A model that randomly targets students (Blind Campaign).

The columns assume different percentages of your total student population that you are able to cover in the campaign, while the rows provide the percentage of your school’s actual high-risk students that will be identified in the campaign based on each model.

The bottom row highlights the substantial relative percentage gains achieved in going from the simple GPA Model to your advanced Student Success Predictive Model and demonstrates that your model is much better at distinguishing between students who are on track to graduate and those that need intervention in order to succeed.

Model Comparison by “At Risk” Students Identified

(For this model, we have assumed that the Total “At Risk” students account for 12.5% of the entire population)
<table>
<thead>
<tr>
<th>Model</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal ball</td>
<td>40%</td>
<td>81%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Your Model</td>
<td>28%</td>
<td>45%</td>
<td>69%</td>
<td>89%</td>
</tr>
<tr>
<td>GPA Model</td>
<td>25%</td>
<td>35%</td>
<td>54%</td>
<td>78%</td>
</tr>
<tr>
<td>Blind campaign</td>
<td>5%</td>
<td>10%</td>
<td>25%</td>
<td>50%</td>
</tr>
</tbody>
</table>

**Relative Percentage Gain**  
12% 29% 28% 14%

So, if your campaign were to select 5% of all students at your institution, a “Crystal Ball Model” would ensure each student was high risk, but you would still only have identified 40% of the students who are actually at risk, because in this scenario we have set your risk threshold for this report at 12.5%.

It is not until the total size of the campaign goes over 12.5%, that we see the Crystal Ball model select 100% of at risk students.

The relative percentage gains compare the “GPA Model” to “Your Model”. Since we are designing this model for strategic outreach, we are most interested in the relative percentage gain at the 10% and 25% points, to see how effective the model would be at identifying the most at risk students compared to the GPA Model.

Relative Percentage Gains is calculated by percentage increases between the two ((69-54)/54=28%).

Your SSPM is high-performing; it can be used confidently to both assess individual students and efficiently design effective, targeted intervention campaigns.

**Introduction**

**Overview**

This document provides information about your institution’s custom Student Success Predictive Model (SSPM). It describes how the model was built; details the top success indicators or “predictors” used in the model and provides metrics characterizing the predictive power of the model.

The SSPM uses your school’s student records to predict the likelihood that any chosen student will Persist to the next term of the regularly scheduled academic year (or graduate before then). This is done by first “training” a statistical model using the records of historical students in order to determine—and assign values to—the items derived from those records that are “predictors” of persistence outcomes.

The model outputs a success score between zero and one estimating the probability that a selected student will persist to the next term. That is, each student’s success score corresponds to the model’s estimate of their likelihood of persisting to the next term. Since it is not possible to build a perfectly
A prescient model, it is important to state that a score of one does not guarantee a student’s persistence. Nor does a score of zero guarantee their failure. A success score of 0.7 for instance, may be interpreted as our expectation that, on average, seven of ten students with this score will persist to the next term.

**Methodology**

The SSPM uses the latest advances in data science to estimate persistence likelihood for each student, from incoming freshmen to nearly-graduating seniors. A customized set of predictors are constructed from student records, and then combined and weighted using an automated training process. EAB’s Data Science team customizes this process for each member, and uses a variety of optimization tools to ensure the best possible performance given the data available.

As described above, the model is trained from recent historical student records; in particular, students satisfying the following criteria were used:

- Matriculated between 2005-08-20 and 2010-01-17.
- Had at least one registered term.
- Were seeking a baccalaureate degree.

Technical details: The model is a combination of several penalized logistic regression models applied to different subgroups of students. The predictors include simple lookups of student records (e.g., high school GPA), as well as composite attributes derived from them whose details are proprietary.

**Your Institution’s Model**

The SSPM includes a wide variety of success indicators called “predictors” in order to ensure maximal predictive power. We use your institution’s historical data to determine the best set of predictors that most accurately reflects the underlying patterns of your students. The items below were found to be good predictors for your institution. The predictors in these lists are not equally important and may not be the same for all subgroups; the statistical model learns how to identify and assign values to the best predictors for each subgroup of students. For instance, we might expect high school GPA to be highly relevant for freshmen, but minimally important for seniors.

**Your Predictors**

The lists below rank the top ten predictors for each subgroup of students included in the model. Please note that transfer credits are incorporated in credit bin determinations.

- Students with 0 Accumulated Credits

**Credits Attempted Current Term**

- Average Credits Attempted per Term
- Age at First Term
- International Indicator
- Estimated Skills
Average Success Outcome of Students Declared in Same Major
Veteran Indicator
Gender
Transfer Indicator

- **Students with Between 1-60 Accumulated Credits**

**Number of Completed Terms**

- Proportion of Transfer Credits
- Total Number of D/F Grades Earned
- Term GPA Earned During First Term
- Average Credits Attempted per Term
- Cumulative GPA
- Average Success Outcome of Students Declared in Same Major
- Estimated Skills
- In State Resident Indicator
- Credits Attempted Current Term

- **Students with Between 61-120 Accumulated Credits**

**Number of Completed Terms**

- Total Number of D/F Grades Earned
- Proportion of Transfer Credits
- Average Credits Attempted per Term
- Average Success Outcome of Students Declared in Same Major
- A student’s cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.
- Ratio of Earned to Attempted Credits
- High School GPA
- Cumulative GPA
- Total Number of W Grades Earned

- **Students with More Than 120 Accumulated Credits**

**Proportion of Transfer Credits**

- A student’s cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.
Cumulative GPA
Major-Skill Alignment
Average Success Outcome of Students Declared in Same Major
Total Number of D/F Grades Earned
Admit Code
Term GPA Earned During First Term
Average Credits Attempted per Term
First Term Transfer Credits

Model Performance

Your SSPM is well-calibrated and its performance has been thoroughly characterized using a “test set” of your historical students that was set aside from the training set and compared to other notional models. These notional models include a fictitious, perfectly prescient Crystal Ball; a GPA Model based exclusively on students’ cumulative GPAs; and a Blind Campaign model that randomly targets students. This section describes the most insightful performance benchmarks and compares your SSPM to these other notional models.

Calibration
Calibration offers an intuitive way to evaluate a model by capturing how close its estimated probability scores are to reality.

Students are divided into different bins along the horizontal axis according to their risk score (to be renamed success score), while the vertical height of each bin indicates the actual persistence rate of the historical students it contains. The horizontal line shows the overall percentage of students that persisted to the next term.
High-Risk Student Identification Rate

The SSPM enables you to rank students by order of risk (i.e., success scores from low to high) so that you can most efficiently target as many high-risk students for intervention as your institution or office/department can effectively handle. Let’s assume, for instance, that you are designing an intervention campaign targeting high-risk students and have the capacity to advise N students. Let’s assume you use different predictive models to generate lists of N targeted high-risk students, and step forward in time to compare their performance by evaluating the percentage of those N students that did not persist to the next term. This performance comparison is summarized in the high-risk student identification rate chart below, which shows the percentage of actual high-risk historical students (i.e., students that did not persist to the next term) identified by the model vs. the percentage of the total student population targeted in the campaign. For example, if you design a campaign that includes 25% of your total student population, then the percentage of your school’s high-risk students identified by the campaign will be 100, 69, 54, and 25 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

High-risk student identification rate provides a powerful and transparent performance benchmark of model performance; the large performance enhancement gained in going from the simple GPA Model to your advanced SSPM is clearly visible in the chart.
High-risk student identification rates can also be converted to actual numbers of students and compared across different accumulated credit subgroups, as shown in the figure below for campaigns targeting 25% of the total student population.
Lift
We may divide the percentage of actual high-risk students identified by a given model by the percentage found by a Blind Campaign to create a new metric called “lift”. For instance, a lift of two would mean that a campaign based on your SSPM identified twice as many high-risk students as a Blind Campaign, while a lift value less than one would indicate that your model identified fewer actual high-risk students than simply choosing from your student population at random. Considering a campaign that includes 25% of your total student population, if we compare to the blind campaign, lift is 7.05, 2.77, 2.21, and 1.00 for the Crystal Ball, your SSPM, GPA Model, and a Blind Campaign, respectively.

Separation
Displaying the distributions of success scores for students in the historical test set who did and did not persist to the next term also provides an intuitive sense of a model’s performance. We see in the charts below that successful students (light gray) typically have higher scores than unsuccessful ones (dark gray) for both your SSPM and the GPA Model, as you would expect, but that your SSPM is much better at separating these two student populations from each other. That is, the graphic demonstrates that your SSPM ascribes high success scores to successful students and low success scores to unsuccessful students more accurately than the GPA Model. A perfect prediction would result in complete separation between the students (shown in the Crystal Ball chart on the right).

Conclusion
The performance of your institution’s Student Success Predictive Model has been extensively optimized and evaluated; the model will provide your school and its advisors with invaluable and otherwise unobtainable insight into your students’ likelihood of academic success. The model incorporates the latest breakthroughs in statistics and data science and places your institution at the cutting edge of student-insight technology. Your advisors may use it with confidence to both assess individual students and design effective and efficient targeted campaigns.
Appendix N1. Evaluating AUC

We commonly use AUC to measure and tune the performance of your Student Success Predictive Model across your institution’s entire student population and different subgroups. AUC stands for Area Under the Curve and is a measure used extensively in data science, which ranges from 0.5 (pure chance) to 1.0 (Crystal Ball). We evaluate your SSPM’s AUC in comparison to the notional GPA Model; your SSPM’s larger AUC indicates that it identifies high-risk students more accurately than the GPA Model. This is the type of rule-of-thumb based approach that academic advisors intuitively know is useful.

The table below shows AUC values for your SSPM and the GPA Model.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA Model</td>
<td>0.75</td>
</tr>
<tr>
<td>Your Model</td>
<td>0.84</td>
</tr>
</tbody>
</table>

As part of validating your SSPM, we examine subgroups of students to ensure that it consistently performs. The figures below show the AUC values for students with different levels of accumulated credits and for Transfer/Non-Transfer students.

AUC for Students with Different Numbers of Accumulated Credits
Appendix N2: High-Risk Student Identification Rate for Murky Middle and Top Performing Students

Your Student Success Predictive Model’s performance varies across different subgroups of students. This appendix provides plots and tables evaluating model performance in terms of high-risk student identification rate for two student subgroups: Murky Middle and Top Performing. The same plots provided for the overall student population in the main body are shown in this appendix for two student subgroups.

Murky Middle
Murky Middle students are defined as those students whose cumulative GPAs are between 2.0 and 3.0.
Model Comparison by “At Risk” Students Identified – Murky Middle

*(For this model, we have assumed that the Total “At Risk” students account for 12.5% of the entire population)*

<table>
<thead>
<tr>
<th>Sample % of Population Selected in Campaign</th>
<th>Crystal ball</th>
<th>Your Model</th>
<th>GPA Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>40%</td>
<td>19%</td>
<td>12%</td>
</tr>
<tr>
<td>10%</td>
<td>80%</td>
<td>36%</td>
<td>22%</td>
</tr>
<tr>
<td>25%</td>
<td>100%</td>
<td>64%</td>
<td>46%</td>
</tr>
<tr>
<td>50%</td>
<td>100%</td>
<td>87%</td>
<td>76%</td>
</tr>
</tbody>
</table>

**Relative Percentage Gain**  58%  64%  39%  14%

Top performing students are defined as those students whose cumulative GPAs are greater than 3.
Model Comparison by “At Risk” Students Identified – Top Performing Students

(For this model, we have assumed that the Total “At Risk” students account for 5.6% of the entire population)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sample % of Population Selected in Campaign</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>Crystal ball</td>
<td>88%</td>
</tr>
<tr>
<td>Your Model</td>
<td>30%</td>
</tr>
<tr>
<td>GPA Model</td>
<td>9%</td>
</tr>
<tr>
<td>Blind campaign</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Relative Percentage Gain</strong></td>
<td><strong>233%</strong></td>
</tr>
</tbody>
</table>

Appendix N3 – Predictor Descriptions

The list below provides detailed descriptions of all the predictors used in your model. We discussed the most important among these in the “Your Predictors” section of the report. This list is ordered alphabetically.

- A student’s cumulative GPA ranked in terms of percentile when compared to other students declared in the same major.: A student’s cumulative GPA ranked in terms of percentile when
compared to other students declared in the same major. This percentile score ranks students in comparison to the performance of their peers’ in the same major; e.g., a sociology student with a score of 80 has a higher cumulative GPA than 80% of all students declared in the sociology major. Students declared in multiple majors are assigned a percentile value that corresponds to the mean average of their scores for each major.

- Admit Code: A student’s admission type (i.e., first time freshman, first time transfer, conditional admit, etc.)
- Age at First Term: A student’s age upon starting their first term at your institution.
- Average Credits Attempted per Term: The average number of credits a student has attempted per term.
- Average Success Outcome of Students Declared in Same Major: This score indicates the average success outcome of all students enrolled in a given student’s chosen major. E.g., if the model’s success outcome is whether a student eventually graduates, and 90% of chemistry students do, then the score will be 90% for all chemistry students. Students declared in multiple majors, however, are assigned the mean average score across all of their majors.
- Credits Attempted Current Term: The number of credits a student is attempting in the current regular term. (The number of credits a student attempted in the most recent regular term is used in the case that a regular term is not currently in session.)
- Cumulative GPA: A student’s cumulative GPA.
- Estimated Skills: A student’s estimated academic skills. More specifically, we identify underlying patterns in the grades students earn in different courses – e.g., some students may have a history of excelling in math-related courses but not writing-related courses – and call the discrete factors behind these patterns “skills”.
- First Generation Indicator: “Yes” or “No” indicator of whether any of an individual’s parents have ever earned a bachelor’s degree.
- First Term Transfer Credits: The number of credits a student transferred from other institutions upon matriculation.
- Gender: A student’s gender.
- Grade Variance in Previous Term: The variance in the grades a student earned in their previous term.
- High School GPA: A student’s high school GPA.
- High School Percentile: A student’s high school rank in terms of percentile.
- High School Size: The size of an individual’s high school student body.
• In State Resident Indicator: A “Yes” or “No” indicator of whether a student is a resident of your institution’s home state.

• International Indicator: “Yes” or “No” indicator of whether an individual is an international student.

• Major-Skill Alignment: A proprietary measure of how well a student’s current major(s) are aligned with their previously demonstrated academic “skills”. More specifically, we identify underlying patterns in the grades students earn in different courses – e.g., some students may have a history of excelling in math-related courses but not writing-related courses – and call the discrete factors behind these patterns “skills”. We then construct a metric gauging how well a student’s demonstrated skills align with those of their chosen major(s).

• Number of Completed Terms: The number of terms a student has completed at your institution.

• Number of D/F Grades Earned in Previous Term: The number of D and F grades a student earned in their previous term.

• Number of W Grades Earned in Previous Term: The number of courses a student has withdrawn from in their previous term.

• Overall Grade Variance: The overall variance in the grades a student has earned from your institution.

• Proportion of Transfer Credits: The proportion of a student’s credits that were earned at another institution.

• Ratio of Credits Attempted Current Term to Prior Term: The number of credits a student attempted in the current regular term as compared to the number of credits they attempted in the prior regular term. (The most recent regular term and the one prior to it are used in the ratio in the case that a regular term is not currently in session.)

• Ratio of Earned to Attempted Credits: The overall number of credits a student has earned divided by the number of credits they have attempted.

• Recent Change in GPA: The difference in a student’s GPA from the prior two complete terms.

• SAT/ACT Math Score Percentile: A student’s highest percentile achieved in either the SAT or ACT math test. We calculate a student’s math percentile as the highest percentile they earned in either the SAT or ACT math tests. A percentile score ranks students in comparison to their peers’ performance; e.g., a percentile score of 80 indicates that a student outperformed 80% of his peers in either the SAT or ACT math tests.

• SAT/ACT Verbal Score Percentile: A student’s highest percentile achieved in either the SAT or ACT verbal test. We calculate a student’s verbal percentile as the highest percentile they earned in either the SAT or ACT verbal tests. A percentile score ranks students in comparison to their peers’
performance; e.g., a percentile score of 80 indicates that a student outperformed 80% of his peers in either the SAT or ACT verbal tests.

- Term GPA Earned During First Term: The term GPA a student earned in their first term at your institution (transfer courses are excluded).
- Total Number of D/F Grades Earned: The total number of D and F grades a student has earned at your institution.
- Total Number of W Grades Earned: The total number of courses a student has withdrawn from at your institution.
- Transfer Indicator: “Yes” or “No” indicator of whether the student transferred from another institution.
- Trend in Number of D/F Grades Earned per Term: A measure of how many D and F grades a student has recently earned relative to past performance.
- Trend in Number of W Grades Earned per Term: A measure of how many courses a student has withdrawn from recently relative to past performance.
- Trend in Term GPA: A measure of the change over time in a student’s term GPAs.
- Veteran Indicator: “Yes” or “No” indicator of whether a student is a veteran of the United States Armed Forces.