CCRC COMMUNITY COLLEGE RESEARCH CENTER

TEACHERS COLLEGE, COLUMBIA UNIVERSITY

Practitioner Perspectives on the Use of Predictive Analytics in Targeted Advising for College Students

Serena Klempin
Markeisha Grant
Marisol Ramos

May 2018

CCRC Working Paper No. 103

Address correspondence to:

Serena Klempin
Research Associate, Community College Research Center
Teachers College, Columbia University
525 West 120th Street, Box 174
New York, NY 10027
212-678-3091
Email: sck2130@tc.columbia.edu

Funding for this study was provided by the Bill & Melinda Gates Foundation and the Leona M. and Harry B. Helmsley Charitable Foundation. The authors wish to thank Melinda Karp, Drew Allen, Elisabeth Barnett, Thomas Bailey, and Laura Gambino for reviewing previous drafts of the paper and providing invaluable feedback. The authors also wish to thank Doug Slater and Kimberly Morse for their excellent editorial guidance and revisions.

Abstract

This study examines the perspectives of college personnel engaged in the consideration, launch, and use of predictive analytics tools for targeted advising. Based on interview data from college staff members at nine public two- and four-year institutions at different stages of implementation of predictive analytics software, we provide a broad look at the positive and negative reactions to this potentially influential educational innovation. Our findings indicate that, overall, advisors and other end users of predictive analytics tools were more critical of them than either administrators or college personnel who helped develop their use as part of a campus-wide advising reform. Our findings also show that, compared with personnel at colleges in the early stages of planning and implementing predictive analytics tools, personnel at colleges that had been actively using the tools for some time reported more concerns. While the use of predictive analytics holds substantial promise in helping to target student services, results of the study suggest that a greater examination of concerns about validity, interpretation, and ethics is warranted.

Table of Contents

1. Introduction

Predictive analytics used in the service of enhanced student advising at higher education institutions has begun to garner a great deal of attention over the past few years (Alamuddin, Brown, & Kurzweil, 2016; Ekowo, 2015; Treaster, 2017). More than two dozen vendors are now providing or adapting software for use in several hundred colleges to assist in activities that aim to increase student success (Bryant, Seaman, Java, & Martin, 2017). Typically, the predictive analytics software uses performance, demographic, and self-survey data on students to identify which students are likely to struggle academically and where and how they might do so. This information may then be used by advisors to target students who need particular kinds of help. And while assisting advisors in their efforts with additional information and analysis about student advisees is a primary purpose of the software (and what we focus on in this report), associated vendor tools are typically designed to carry out other functions as well, such as suggesting to students that they might want to consider particular programs of study or courses, and providing information to colleges to help assess the impact of particular courses or interventions.

In press accounts and other literature, predictive analytics for targeted advising has often been presented as a "game changer" that has the potential to help colleges prioritize and improve their student success efforts. And it is perhaps not surprising that much of the early conversation around the use of predictive analytics has focused on its promising benefits rather than on potential downsides or unintended consequences. The study we report on here explores on-the-ground perspectives and opinions of higher education practitioners engaged in the consideration, launch, and use of predictive analytics tools for targeted advising. Based on interview data from college staff members at nine institutions at different stages of implementation of predictive analytics tools, we provide a broad look at the positive and negative reactions to this promising and potentially influential educational innovation.

Our findings indicate that, overall, advisors and other end users of predictive analytics tools were more critical of them than either administrators or college personnel who helped develop their use as part of a campus-wide technology-mediated advising reform. Compared with personnel at colleges in the early stages of planning and implementing predictive analytics tools, personnel at colleges that had been actively using the tools reported more concerns. While the tools hold substantial promise, results of the study suggest that a greater examination of concerns about validity, interpretation, and ethics is warranted. Planning for the adoption of particular software and tools with these considerations in mind will help ensure that the use of predictive analytics for targeted advising at colleges provides a valuable service to students.

2. Background and Motivation for the Study

Predictive analytics refers to the use of highly sophisticated statistical algorithms that sift through large amounts of assorted historical data to identify notable patterns and thus predict future related outcomes. Associated with other terms such as "data mining," "predictive modeling," "big data," "data science," and "machine learning," predictive analytics represents a powerful branch of computer and statistical science with myriad applications. Interactive software that makes use of predictive analytics has been developed in a wide array of sectors to carry out a broad range of functions, including sales forecasting, fraud detection, and improved manufacturing. Predictive analytics is increasingly used in higher education as well for such disparate purposes as offering adaptive learning courseware that modifies content for individual students based on their interaction with the system, predetermining which student loan borrowers may be at risk of default, and forecasting enrollment patterns by rating prospective students' likelihood of attending (Beckwith, 2016; EAB, 2017; Ekowo & Palmer, 2016; Fain, 2015). In terms of its use in relation to targeted advising, predictive analytics software is at its core a system that uses data on current and past students to evaluate risks concerning the progression and completion of current enrollees.

2.1 How Predictive Analytics Is Used in Targeted Advising

Although some predictive analytics tools are developed by colleges themselves, most institutions hire an external technology vendor to facilitate adoption of one. Some vendors partner with institutions to create specific approaches, algorithmic formulas, and visualizations of data and results, but others rely on proprietary models developed with little or no input from colleges (Ekowo, & Palmer, 2016, 2017). Despite variations in the development of these predictive analytics systems, the student data that are analyzed by them typically fall into one of five categories: (1) preenrollment data (e.g., demographic information, high school GPA), (2) academic data (performance in college courses), (3) noncognitive data collected from student surveys administered by the college (e.g., self-reporting about motivation and self-efficacy), (4) data on service use (e.g., tutoring), and (5) engagement data (e.g., attendance at campus events) (Burke, Parnell, Wesaw, & Kruger, 2017).

One key task of a predictive analytics system is to use large swaths of historical and current data to identify which students are likely to struggle academically and drop out. The goal is then to connect those students to support services before they fall too far off track. For example, the University of North Carolina at Greensboro uses a predictive analytics model containing 80 variables to rank students' likelihood of persisting, and then divides students into ten categories based on level of risk. The variables that are used relate to a variety of risk factors concerning academic preparation (in which standardized test scores and high school attended are included as variables), academic behavior (in which class attendance and use of tutoring services are included), and socio-emotional stressors (in which even distance from home is included). The model can be run as frequently as weekly when new information becomes available. Based on the analysis, student success staff conduct targeted outreach to intervene with those students identified as at greatest risk of not completing college (June, 2017).

It is important to recognize that vendor tools can perform a number of other functions. A system with a student interface may provide alerts to students, or it may even direct students toward courses and majors in which they are predicted to be most likely to succeed. For example, the Major Matcher software feature at Georgia State University (GSU) allows undergraduate students to evaluate their probability of success in every major offered by the university based on the performance of previous students as well as the students' own past grades (Georgia State University, 2014).

2.2 Discourse on Predictive Analytics and Advising in the Media and Elsewhere

Much of the media coverage on the use of predictive analytics tools in higher education, as in other sectors, has been very positive. For example, the large-scale successful use of predictive analytics by GSU is a frequent focus of press coverage and is emblematic of the potential of predictive analytics to help with advising and student success improvement efforts.

After implementing a range of reforms, including a predictive early warning system coupled with advising redesign, GSU increased its graduation rates and closed the completion gap for students of color and for low-income and first-generation students (Georgia State University, 2014; Renick, 2016). The predictive early warning system analyzes ten years' worth of GSU student data and identifies over 800 risk factors—including indicators such as registering for an unnecessary class or doing poorly in a class that is a prerequisite for subsequent coursework—that trigger automatic alerts to advisors (Renick, 2016). At the same time as it launched its predictive analytics system, the university hired 42 new advisors to respond to alerts, bringing the student-to-advisor ratio down from over 700 to 1 to 300 to 1 (Georgia State University, 2014). In addition, the predictive analytics system assesses each student's academic performance and rates their ability in three core functional areas: math and science, humanities and art, and business and policy. Advisors are then able to use this information to guide discussions about students' academic paths (Georgia State University, 2014).

Not surprisingly, media coverage of GSU has focused on the impressive outcomes it has achieved. These outcomes include a 4 percent increase in retention, a 5 percentage point increase in the four-year graduation rate, a 6 percentage point increase in the six-year graduation rate, and $10 million in additional revenue from tuition and fees as a result of the increase in retention (Dimeo, 2017; Treaster, 2017). Similarly positive coverage, both of GSU and other universities, has appeared in a number of other sources, including NPR (Kamenetz, 2016), *EdSurge* (McNeal, 2016), the *Chronicle of Higher Education* (Hoover, 2015; June, 2017), and the *New York Times* (Steinhauer, 2017), among others. In addition, vendors have begun to release reports highlighting success stories, primarily of increases in retention and completion, such as a recent brief by the Education Advisory Board (EAB) profiling ten colleges and universities (EAB, 2017).

Nonetheless, as with any new innovation, widespread enthusiasm has also been tempered by concerns about unintended consequences and potential problems. Some of these concerns focus on using historical data on previous students—which may include demographic data such as race/ethnicity and gender—to predict outcomes for current individual students. For example, a recent piece in the *Hechinger Report* highlighted at-risk students' concerns about being labeled as failures before they have been given a chance to succeed (Dobo, 2017). In his *Inside Higher Ed* column, Reed (2017) similarly worried that applying predictive analytics to particular students (rather than groups) might inadvertently produce a "stereotype threat" in which students targeted as needing help might be nagged by the thought that "people like me aren't good at x." And Warner (2016) stressed the importance of engaging with students as individuals in order to truly understand why they might be at risk rather than relying on data to determine when to intervene. Others have cautioned that blind faith in predictive analytics may cause institutions to overlook the challenges associated with interpreting the data appropriately and have emphasized the need for training on how to apply predictive analytics in practice (Phillips & Horowitz, 2017; Salisbury, 2017). Most dramatically, the alleged attempt by the former president of Mount St. Mary's University to use data to identify new freshman students most likely to drop out and to encourage them to do so before they were included in the retention numbers used in college rankings drew public condemnation and illustrated the potential for not just misuse, but abuse, of predictive analytics (Johnson, 2017).

Citing similar concerns as those sketched out above, some research and policy organizations have begun calling for a more cautious approach to implementing predictive analytics in higher education, particularly in advising. For example, a recent report by the New America Foundation highlighted several potential ethical problems centered around reinforcement of implicit biases, discouragement of students from pursuing their interests, and violation of student privacy (Ekowo & Palmer, 2017). Summarizing concerns raised in a number of different sources, a report by Ithaka S+R addressed many of the same issues (Slade, 2016), and it also noted reservations about technical aspects of predictive analytics, including the accuracy of predictions and the quality of the data used, and about practical aspects, such as ensuring that staff and faculty have the time and knowledge to effectively respond to the results of the analysis. The study reported on here was undertaken in part to introduce further perspectives on the use of predictive analytics for enhanced advising by providing additional voices—those of practitioners using the tools or contemplating or planning their use in nine colleges across the country.

3. Data and Method

Data used in this paper are drawn from a larger qualitative study project about the implementation of technology-mediated advising reforms among colleges participating in the Integrated Planning and Advising for Student Success (iPASS) grant initiative, funded by the Bill & Melinda Gates Foundation and the Leona M. and Harry B. Helmsley Charitable Trust. Data for the paper consist of interviews with key personnel at a subsample of iPASS grantee colleges, including 3 public four-year colleges and 6 public two-year colleges, that have been engaged with predictive analytics as part of their work for the grant. Interview questions for the larger qualitative study were designed to cover a broad range of implementation issues related to several different types of advising technologies (e.g., education planning and early alerts tools) and were not intended to evaluate colleges' use of predictive analytics or to assess the effectiveness of predictive analytics tools. Thus, questions did not, for example, ask for detailed information about the variables used in predictive analytics algorithms. Instead, the data provide insight into practitioners' general opinions based on their knowledge of and experience with the application of predictive analytics to advising.

3.1 Site Selection

Data collection for the study occurred between March 2016 and February 2017. Because the colleges were at different stages of considering or implementing predictive analytics at the time of the study—some of the colleges were just beginning to think about investing in predictive analytics, while one of the colleges had been actively using a predictive analytics tool for over a year—some practitioners' opinions were based primarily on their conceptual knowledge of predictive analytics, while others' opinions were based on their experience of having used predictive analytics. To distinguish between these perspectives, we created three categories for classifying implementation progress, based on project descriptions contained in grant proposals, first-year grant reports, and interview data:

- *Considering*: A few individuals at the institution were interested in pursuing predictive analytics, but predictive analytics tools had not yet been purchased at time of the site visit.

- *Planning*: Predictive analytics tools had been purchased and were in the process of being set up and tested, but they were not yet fully implemented institution-wide.
- *Using*: Predictive analytics tools had been implemented institution-wide.

Of the nine institutions, we categorized 3 two-year colleges as being in the *considering* stage, 1 four-year college and 2 two-year colleges as being in the *planning* stage, and 2 four-year colleges and 1 two-year college as being in the *using* stage.

The predictive analytics tools used by the nine institutions were provided by three distinct vendors but nonetheless carry out similar functions that enable colleges to identify at-risk students based on individual risk scores. The colleges that were further along in planning for or implementing their use of predictive analytics were also using the resultant information to devise strategies for targeting groups of students that appeared to be at higher risk of not completing their programs of study based on the data. In addition, several of the colleges were integrating or planning to integrate data used in their predictive analytics software with data gained from other related tools in order to provide a more comprehensive picture of risk. For example, one of the four-year universities actively using predictive analytics was working on combining predictive risk scores with information about course-specific performance and behaviors based on flags raised in early alert systems as well as information about projected time to degree from degree audits. Finally, the other four-year university actively using predictive analytics was doing so, in part, to identify the courses most predictive of success in particular majors, and to predict the courses in which individual students are likely to do well.

3.2 Participant Sample

In order to include a range of perspectives, we purposefully selected study interviewees (participants) at each site who represented three distinct functional roles: administrators, core team members directly involved in the reforms, and end users. Administrators (e.g., presidents, provosts, deans) provided an overview of college structures, strategic priorities, and reform goals. Core team members were responsible for launching technologies and overseeing other changes; these individuals represented a diverse array of departments involved in technology-mediated advising reform (e.g., staff members from academic affairs, advising and other student support services, institutional research, and IT). Finally, end users (e.g., advisors, faculty members, and other student services personnel) were the people asked to integrate the technologies and related reforms into their daily practices.

Analysis focused on a subset of participants within the nine colleges, each of whom expressed a clear opinion about the use of predictive analytics for targeted advising ($n = 58$).[1] Participants were evenly divided between two-year colleges ($n = 29$), and four-year colleges ($n = 29$). They included 12 administrators, 20 core team members, and 26 end users.[2] The vast majority of participants came from institutions that were either in the planning stage of implementing predictive analytics ($n = 26$) or actively using predictive analytics ($n = 25$). Only 7 participants came from institutions that were in the considering stage.

3.3 Data Analysis

Data analysis occurred in two phases. During the first phase, we developed a broad coding scheme related to overarching research questions about the implementation of technology-mediated advising reform (that we make use of in this and related studies). The coding scheme covered a variety of topics related to technology and advising, including a general code to capture any mention of predictive analytics being used for targeted advising, as described by Ekowo and Palmer (2016). After noticing that discussions about the benefits and challenges of using predictive analytics were prevalent throughout the data, we decided to conduct a second phase of analysis to explore participants' opinions about predictive analytics in greater detail. Taking an inductive approach to the analysis, we used impressions from our initial review of the data to guide the development of new predictive analytics subcodes (Braun & Clarke, 2006). To begin the process of recoding, we first created two categories pertaining to opinions about predictive analytics: reasons for being critical of predictive analytics and reasons for being supportive.

[1] A "clear opinion" was defined as a statement that (1) demonstrated knowledge of planned or actual use of predictive analytics at the institution and (2) expressed a readily classifiable reason for having a positive (supportive), negative (critical), or mixed reaction about the use of predictive analytics.
[2] In addition to individual interviews, from which we gained most of our data, we also conducted two focus groups. Because the transcripts do not differentiate between individuals in the focus groups, each focus group was counted as a single case or "person."

Nearly all the participants who discussed the benefits of predictive analytics described the same broad reason for being supportive—that a better understanding of student risk enables the college to target advising and other supports more effectively. Another benefit that was mentioned by three participants pertained to course recommendations. Additionally, two other participants each mentioned a reason for being supportive of predictive analytics that was related to the benefit of adopting predictive analytics for the institution. Thus we ultimately created three subcodes for reasons for being supportive (the first of which dominates in our findings):

Reasons for supportive opinion (benefits):

- Enables more effective advising
- Provides course recommendations that set students up for success
- Benefits the institution

Because a significant number of participants raised several different concerns repeatedly, we created multiple additional subcodes related to reasons for expressing a critical opinion. As we recoded the data, we modified these subcodes to more accurately reflect our refined interpretations of the data. The final subcodes included:

Reasons for critical opinion (concerns):

- Concern about the validity of variables used as measures of risk
- Concern about the validity of predicted outcomes
- Lack of understanding about how predictions are calculated
- Insufficient opportunities for training/professional development in using the tools
- Concern about depersonalizing effects on the relationship between advisor and advisee
- Concern about providing legitimacy for racial/ethnic bias
- Concern about "predetermining" failure/counseling students out

Finally, because several of the reasons for being critical of predictive analytics seemed to be addressing similar core issues, we grouped the reasons for being critical into three sub-categories:

- *Validity*: Concerns about the validity of predictive analytics as an assessment of student risk. This category includes concerns about (a) the validity of variables included as measures of risk, and (b) the validity of the predicted outcome as a true indicator of a student's overall likelihood of being at risk.
- *Interpretation*: Concerns about the inability to interpret inputs into the algorithms and the difficulty of interpreting the results generated by the algorithms. These are related to (a) frustration over not having received adequate explanation about what information is contained in the algorithms and how that information is weighted in the algorithms and (b) the expressed need for professional development concerning how to use predictive analytics to inform interactions with students and, in particular, how to communicate "risk" to students.
- *Ethics*: Concerns about whether predictive analytics may negatively impact the relationship between advisors and students or harm students. These are related to (a) the possibility that discussing risks as identified by the algorithms dilutes or impairs the personal relationship between advisors and students in part by devaluing advisors' ability to assess students as individuals, (b) the possibility that algorithmic results serve to legitimate or strengthen racial/ethnic biases among advisors, students, and others, and (c) that the results may encourage the counseling of students out of majors or college altogether.

In addition to wanting to understand distinct reasons why participants were either supportive or critical of predictive analytics, we also wanted to understand how participants made sense of these various reasons to formulate an overall point of view. It quickly became apparent that many participants were attentive to both concerns and benefits (or potential concerns and benefits) of predictive analytics. In order to measure how prevalent particular concerns and benefits were, we calculated frequencies based on the number of individuals who mentioned a concern or benefit, not the number of times a concern or benefit was mentioned. If a participant mentioned the same concern or benefit multiple times, it was counted once. If a participant mentioned more than one concern, each unique concern was counted separately. None of the participants who discussed reasons for being supportive of predictive analytics mentioned more than one benefit. Finally, we examined all of the clearly opinionated statements each individual made about predictive analytics and summarized his or her point of view using three categories:

- *Supportive*: All statements by participant indicated support for predictive analytics.
- *Critical*: All statements made by participant were critical of predictive analytics.
- *Mixed*: Participant acknowledged potential benefits of predictive analytics while also raising concerns.

After we recoded and categorized the data, we used various analytic tools in Dedoose and Excel spreadsheets to examine variation in participants' opinions about predictive analytics based on the college's stage of implementation, as well as on participant type (administrator, core team member, end user). Because most of the information in the literature about how colleges are using predictive analytics comes from four-year colleges and little has been reported based on the experiences of two-year colleges, we also analyzed the data by sector in order to look for variation in points of view between participants from two- and four-year colleges.

4. Findings

We first report descriptive findings that summarize our classification of participants' points of view about predictive analytics and that explore those points of view in more detail. We then report comparative analytic findings that describe the relationship between point of view on predictive analytics and institutional sector, implementation stage, and participant type.

4.1 Overall Findings

Categorization of participants. Out of the 58 participants who expressed a clear opinion about predictive analytics for targeted advising, the largest number discussed both the pros and cons of predictive analytics (23, or 40 percent). Fourteen (24 percent) were unequivocally positive. Twenty-one (36 percent) were unequivocally critical. Thus, the largest portion of opinionated stakeholders were weighing potential benefits against risks as they considered how to effectively implement and use the technology.

Looking at the total number of individuals who raised a concern about predictive analytics (combining those who were unequivocally critical with those who identified pros as well as cons), 44 participants (76 percent) expressed at least one concern about predictive analytics. Likewise, looking at the total number of individuals who identified a reason for being supportive of predictive analytics (combining those who were unequivocally positive with those who identified pros as well as cons), 37 participants (64 percent) identified a benefit of predictive analytics.

Benefits of using predictive analytics for targeted advising. Although proponents discussed both the benefits of providing targeted support for students identified as being generally at risk of not completing, and of tailoring course recommendations for students identified as being at risk of not doing well in particular courses, the vast majority focused on the benefits of targeted advising for students generally at risk. As mentioned, only two participants discussed reasons for being supportive of predictive analytics in terms of benefits for the institution.

Table 1.
Frequency of Benefits

Benefit	Number of Participants Mentioning Benefit
Enables more effective advising	32
Provides course recommendations that set students up for success	3
Benefits the institution	2

Enables more effective advising. Thirty-two of the 37 participants (86 percent) who expressed some degree of support for predictive analytics (both those who were unequivocally positive and those who discussed pros as well as cons) believed predictive analytics provides a more effective method of using data to proactively identify students at risk, thereby enabling colleges to better target support services and promote academic success. For example, an administrator from a community college considering investing in predictive analytics noted that the advisors on campus had reported being frustrated by a lack of easy access to information about students and had expressed interest in predictive analytics. The advisors were hopeful that predictive analytics would enable them to concentrate on the students for whom additional support would have the biggest impact.

> The other thing they [advisors] were looking for was … is there some way based on the information in Banner about a student where we can predict whether or not they were more likely … of completing. So are there some predictive things and red flags that could show up that would help an advisor take a case load of 400 to 600 students and kind of prioritize them as the more at-risk students, the less at-risk students, and the not-at-risk students? So they could focus their time on the students who probably would benefit most.

Similarly, a core team member from a four-year university in the planning stage of implementing predictive analytics discussed the benefits of being able to use risk scores related to students' likelihood of completing in order to conduct more targeted outreach campaigns.

> It creates the score that you can attach to a student, that basically sort of would inform an advisor that this student has the characteristics … that we associate with a high risk of not persisting.… So if we need to make choices about outreach projects, we want to contact all students that meet these particular characteristics.

Provides course recommendations that set students up for success. Three participants (8 percent) discussed the benefits of being able to identify students at risk of not succeeding in particular courses. These participants viewed this information as critical for directing students toward courses in which they are likely to succeed, and for ensuring that students receive support to promote their success in the courses in which they are likely to struggle. Describing how his university uses predictive analytics in this way, an administrator spoke hypothetically about how he would advise a math major who had just completed the first semester.

> I could pull you up in [the predictive analytics tool] and say, okay, well this semester you took these five classes and you got a C+, B, A, B, and a C+. Pretty good grades for a freshman. Not too bad. But, as a math major, I can see that the class you got a C+ in—really to maximize your chances of success in the next class—you needed that to be a B- not a C+. So that says to me this next class is going to be tough for you. So we are going to put you in that class at your most alert time of the day, we are going to start you in tutoring with that class from day one, and we are not going to give you another really hard class this semester. We are going to take some of the classes that we think are going to be easier for you, so you can really stay focused on that one.

Benefits the institution. Only two of the 37 participants (5 percent) who mentioned a benefit of predictive analytics identified other distinct reasons for being supportive. Both participants described ways in which they hoped the successful adoption of predictive analytics could benefit their institutions. A staff member on the core team at a four-year university in the planning stages of implementing a predictive analytics tool was hopeful that better data about how and where to target services would increase the institution's return on investment by improving retention and completion rates, and a core team member from a two-year college that was also in the planning phase of implementing predictive analytics suggested that predictive analytics would enable the college to better track the impact of interventions by demonstrating the correlation between participation and risk scores.

Concerns about the use of predictive analytics for targeted advising. As described in the section on recoding for predictive analytics, participants identified seven distinct concerns about predictive analytics, organized into three types.

Table 2.
Frequency of Concerns by Type

Concern	Type of Concern	Number of Participants Mentioning Concern
Validity of predicted outcomes	Validity	17
Need for professional development/lack of training	Interpretation	15
Validity of variables used as measures of risk	Validity	9
Lack of understanding about how predictions are calculated	Interpretation	8
Risk of "predetermining" failure/counseling out students	Ethics	7
Risk of depersonalizing the advisor-advisee relationship	Ethics	6
Risk of reifying racial/ethnic bias	Ethics	3

After looking at the number of individual concerns, we also combined concerns of the same type in order to understand the general kinds of issues that were emerging the most frequently. Concerns about validity were the most common, closely followed by concerns related to interpretation.

Table 3.
Frequency of Type of Concern

Type of Concern	Number of Participants Mentioning Type of Concern
Validity	26
Interpretation	23
Ethics	16

Validity. Twenty-six participants had concerns about validity. The most commonly mentioned concern, identified by 17 participants, was a lack of trust in the validity of predicted risk scores. When risk scores did not align with what participants understood about particular students, buy-in and willingness to use predictive analytics appeared to decline dramatically. As the following quote from an advisor at a four-year university actively engaged in using predictive analytics illustrates, the perception that risk scores do not provide a valid indication of students' true ability was frustrating for practitioners being asked to use the data. It made little sense to this advisor why someone with a near-perfect GPA would be placed at high risk, while someone with a low GPA would be placed at low risk.

> [It raises concerns] when you have someone who has a 3.9 and a high risk of not graduating. … Or you have someone with like a 2.1 who has a low chance of not graduating because they have passed the general education courses, but they cannot pass a class in their major. So it is something that [makes one distrust the results]. People say, "Alright, we see high, low, medium, and that's great," but it does not matter.

Illustrating the importance of having valid indicators of risk to produce valid predictions, the advisor speculated that because the student with a low GPA had a low risk of not completing, the algorithm must have weighted criteria other than overall GPA rather heavily and must have considered passing general education courses as more important than passing a class for the major. Without knowing the content of the algorithm, however, it was impossible for him to know which variables carried the most weight, leading him to conclude that the risk score "does not matter."

Other participants (9 in total) expressed more direct concerns about the validity of what was being put *into* predictive analytics.[3] For example, a core team member from a community college using predictive analytics worried that their algorithm relied too heavily on the vendor's standard indicators, such as using the characteristics associated with students who leave college after their first year as risk factors. This core team member argued that students who leave the community college after their first year may be successfully transferring to a four-year institution rather than dropping out of college. Algorithms that treat leaving after the first year as an unsuccessful outcome may therefore be failing to distinguish between these two groups and might be advancing an invalid risk model.

> …The model needs to change for a community college because for us, you know, a student may not come back to us in the fall, but not because they're dropping out, but because they are transferring. Well that is still a successful student. … So I think that they are going to have to adjust their model to truly try to get at the students who are not successful because they're not successful, rather than not completing here because that was never their intention.

Interpretation. Twenty-three participants had concerns related to their ability to interpret and apply predictive analytics data. The second most frequently mentioned concern overall, identified by 15 participants, centered on inadequate training and support regarding how to apply predictive analytics in practice. For example, an advisor from a four-year university in the planning stage of implementing predictive analytics was frustrated because he had not received any training that would enable him to translate a risk factor into an intervention for a specific student.

> [Advisors] haven't really had training. We've seen what the dashboard looks like and things in the criteria, and where [the system] falls short is, it is pointing out the risk factors

[3] The distinction between validity of inputs versus outcomes is a tenuous one, but we retain it because, as we discuss in the subgroup findings below, particular categories of college personnel tended to mention each concern.

> of students, and I'm like, "So how do you address that risk factor?" I'm sitting here with [a student]. How do I address that? ... Over the summer there should be some training.

In addition to seeking more guidance on how to respond to predictive analytics, participants also requested more information about how the algorithm works, and what specific pieces of information are included for analysis. One advisor was so frustrated by the lack of explanation that he mentioned feeling like "my head wants to explode":

> I know it is going to gather data and give us information on students, but I don't understand how it works in the background and how it grabs the different pieces of information that it does. You know, like I don't know, it makes my head wants to explode thinking about it. … I don't understand how it is gathered behind the scenes, so that is a little bit difficult for me.

Without this explanation, the advisor was unclear about how he was supposed to use predictive analytics. After briefly trying to describe his understanding of what he was supposed to do, he concluded, "I guess it's still really kind of confusing to me."

Ethics. Sixteen participants had ethical concerns. Six participants expressed misgivings about the impact of predictive analytics technology on the relationship between advisors and students. They felt like they were being asked to trust technology more than their own judgment. They viewed predictive analytics, in particular individual risk scores, as something that was simultaneously taking away their autonomy and depersonalizing interactions with students. They preferred to work with students directly and use the information they already had accessible to them to figure out the best course of action to take with particular students on a case-by-case basis. This was a prevalent theme in a focus group we conducted with advisors at a four-year university actively involved in using predictive analytics.

> Advisor 1: The software doesn't have common sense.
>
> Advisor 4: I just deal with the person.
>
> Advisor 5: I mean they spend so much time putting all this stuff in, but honestly, just looking at just a few things on

> there … I can tell if they're at risk. … I don't need that thing to tell me high, low, medium. I already know.
>
> Advisor 6: And when it says low and the student comes in and you know they're in distress, it's [still a concern].

The last statement in particular highlights the value advisors placed on personal interactions. Even students with a low risk of not completing may encounter serious challenges and hurdles. Advisors conveyed that they want to address the needs with which the student presents, rather than making assumptions about a student based solely on a risk score.

Seven participants were concerned that labeling students as at-risk would negatively impact not only faculty and staff perceptions about students' abilities but also students' perceptions about themselves. For example, an administrator from a two-year college that was considering implementing predictive analytics worried that using background information such as demographics and high school GPA to label students as at-risk would result in counseling students out of majors preemptively, causing students to internalize a sense of themselves as failures before they had even been given a chance to succeed.

> I mean, geez, you can't take a student that walks in and say, "Okay, here's your demographic background, here's your high school GPA, and here's your program of study that you want to be in. You can't, sorry. You're at-risk, so let's put a piece of tape on you and write that blah, blah, blah." And you get to walk around knowing that you're just about to fail any day.

The hypothetical situation described by the administrator here provides an interesting example of how the use of predictive analytics raises new ethical concerns in college advising. Administrators and advisors could just as easily counsel students out of majors based solely on demographics and high school GPA without referring to a predicted risk score. In that case, however, the ethical concern has to do with recognizing and addressing individual biases or preconceptions. When predictive analytics produce a risk score that confirms existing biases or preconceptions, the decision to counsel students out of majors can be attributed to the seemingly objective algorithm behind the score, without having to acknowledge the rationale behind that decision. Writing about the negative impact of predictive algorithms used to assess people in a variety of different contexts, O'Neil (2016, p. 7) warns that "many poisonous assumptions are camouflaged by math and go largely untested and unquestioned."

Three participants were especially troubled by the implications of predetermining failure in this way for students from racial/ethnic minority groups. A core team member from the same two-year college as the administrator quoted above discussed how even seemingly objective measures such as test scores reflect a history of racial and economic segregation, and expressed a concern that the use of predictive analytics would result in singling out minority students from impoverished school districts.

> We have one of our school systems, one of the three that we serve, that has really flipped in the last twenty years. They've become a minority-majority school division. … We talk about the local devastated economy; they were the hub. They've lost tons of population, closed schools, and as often happens, test results have plummeted. We want to be very cautious in how we frame these conversations so that it doesn't look like any school or any division or any population is singled out.

Summary. Overall, while the concerns about predictive analytics were multifaceted and complex, there was commonality in many of the most prevalent concerns raised across participants and colleges. In order to better understand these concerns, we next looked for patterns related to who was supportive or critical of predictive analytics and why.

4.2 Subgroup Findings

After identifying the reasons why participants were either supportive or critical of predictive analytics, we examined variations in both overall point of view on predictive analytics (supportive, critical, or mixed), and in the particular concerns raised, based on sector (two-year versus four-year), each institution's stage of implementation (as an indication of exposure to predictive analytics), and participant type. The outcomes of these analyses follow.

Variation by sector. The numbers of participants in each point of view category (supportive, critical, and mixed) were nearly evenly split between two-year colleges and four-year colleges. Of the 14 supportive participants, 8 were affiliated with a two-year college while 6 were affiliated with a four-year institution. Of the 21 critical participants, 10 were affiliated with a two-year college and 11 were affiliated with a four-year institution.

Table 4.
Variation in Point of View by Sector

Sector	Supportive	Critical	Mixed	Total
Two-year	8	10	11	29
Four-year	6	11	12	29
Total	14	21	23	58

Following a similar pattern, the two most frequently raised concerns were mentioned by similar numbers of participants from two-year and four-year colleges. Seven participants from two-year colleges and 10 participants from four-year colleges were concerned about the validity of predictions, while 7 participants from two-year colleges and 8 participants from four-year colleges discussed a need for professional development. In fact, all concerns except for 2 were raised by participants at both two-year and four-year colleges. All 6 of the participants who voiced a concern about the risk of depersonalized advisor-advisee relationships came from four-year colleges; they represented each of the three four-year colleges in the sample. And all 3 participants who voiced a concern about racial/ethnic bias came from two-year colleges, representing two separate colleges. Overall, therefore, despite the sample limitations discussed previously, the similarity of the distribution of views about predictive analytics across sector suggests that the opinions raised here are generally not unique to one sector versus the other.

Variation by implementation stage. In order to examine variation based on each institution's progress in implementing predictive analytics, we looked at each stage (considering, planning, using) separately and compared the breakdown of participants with supportive, critical, and mixed views in each one. Among the colleges that were in the *considering* stage of implementation, the majority of the 7 participants (57 percent) expressed both pros and cons. Compared to institutions in the *considering (28%)* and *using (12%)* phases, institutions in the *planning* phase had the highest percentage of unequivocally positive participants (9 out of 26, or 35 percent), although the largest number of participants (11 out of 26, or 42%) held mixed views. Finally, the majority of the participants from the colleges actively using predictive analytics were critical (14 out of 25, or 56 percent). Only 3 out of the 25 (12 percent) were unequivocally positive.

Table 5.
Variation in Point of View by Implementation Stage

Implementation Stage	Supportive	Critical	Mixed	Total
Considering	2	1	4	7
Planning	9	6	11	26
Using	3	14	8	25
Total	14	21	23	58

These findings are consistent with the notion that the more deeply engaged in predictive analytics participants became, the more critical they were. The general trend appears to be one of moving from initial curiosity about what predictive analytics can do, to excitement about the potential benefits as systems are being planned or implemented, to concern or discouragement after participants become actively engaged in using predictive analytics.

Supporting the idea that colleges encounter different challenges at different points of the implementation process, the need for professional development was most pronounced at *considering/planning* colleges (14 out of the 15 voiced concerns came from institutions that were *considering* or *planning*). Staff at the institutions that were in the *using* stage were more concerned about validity. All 17 of the concerns about validity of outcomes were made by participants from institutions that were using predictive analytics.

Variation by participant type. Administrators were the most supportive of predictive analytics: 5 out of 12 (42 percent) were unequivocally positive, while the remainder expressed both pros and cons. None were unequivocally critical. Similarly, only 3 of the 20 core team members were unequivocally critical. However, a smaller percentage of core team members than administrators were unequivocally positive (7 out of 20, or 35 percent). Half expressed both pros and cons (10 out of 20). In contrast, 18 out of 26 end users (69 percent) were critical, while 6 expressed both pros and cons, and only 2 were unequivocally positive.

Table 6.
Variation in Point of View by Participant Type

Participant Type	Supportive	Critical	Mixed	Total
Administrator	5	0	7	12
Core team member	7	3	10	20
End user	2	18	6	26
Total	14	21	23	58

Both of the end users who were positive came from colleges that were in the planning stage—they had heard presentations about the benefits of using predictive analytics but did not know much about them and had not used them. Overall, end users (most of whom were advisors) were far more critical than core team members or administrators, particularly end users from institutions actively engaged in using predictive analytics. Thus it appears that those most closely involved in using predictive analytics (both in terms of institutional role and length of institutional exposure to predictive analytics) are the ones most likely to have concerns about them.

Furthermore, the particular concerns raised by end users reflect the fact that they are the ones most directly tasked with taking results from predictive analytics and applying those results to interventions with students. End users at colleges using predictive analytics were most concerned with (1) validity of predicted outcomes (15 out of these voiced 17 concerns came from end users), (2) lack of understanding about how predictions are calculated (6 out of these 8 concerns came from end users), and (3) risk of depersonalization (5 out of these 6 concerns came from end users). On the other hand, reflecting their role behind the scenes in implementing predictive analytics, administrators and core team members were more concerned about the specific information being put into algorithms (all 9 of the concerns about the validity of the variables used in algorithms came from administrators and core team members).

Summary. Interestingly it appears that sector had little impact on the kinds of opinions voiced by participants. What did seem to matter was exposure—both in terms of overall institutional exposure (stage of implementation) and the extent to which individuals' roles within the institution connected them with the use of predictive analytics tools (participant type). The more closely participants were involved with using predictive analytics on the ground with students, and the more advanced the associated institution was with its implementation, the more critical the participants were about the use of predictive analytics. Although the relatively small sample used in this study may limit the generalizability of the results, the fact that participants from different types of institutions were saying similar things about their predictive analytics tools suggests that the findings may be illustrative of more widely held sentiments about the use of predictive analytics for targeted advising.

5. Conclusion

All of the participants we interviewed were committed to improving student success and were hungry for tools and strategies that would enable them to more effectively target student services. Many believed that predictive analytics could and did fulfill that function. In order to maximize the potential for realizing these benefits, however, it is crucial to pay attention to the concerns raised by the stakeholders doing the work. Importantly, different types of concerns will require different responses.

Concerns about the validity of both the variables used in algorithms and the predicted outcomes call for continued quality control, testing, refinement, and assessment of predictive analytics systems. In their guiding principles for the responsible use of predictive analytics, Ekowo and Palmer (2017) stress the importance of proactively planning for and continually monitoring issues related to the concerns about validity discussed here. They emphasize that the accuracy of predictions depends upon having high quality, comprehensive data, and recommend that institutions select vendors based on their willingness to actively collaborate and be transparent about the processes used in their software. Additionally, Ekowo and Palmer advise colleges to ask external evaluators to test algorithms.

The concerns related to the interpretation of predictive analytics we identified suggest a need to (1) provide trainings that clearly explain how predictive analytics works and what data are included (and to the extent possible, what specific factors are predictive of risk and why), (2) devote extensive time and resources to professional development regarding how to apply the analytic results in practice (how advisors should explain to students why they are concerned and identify which specific action steps should be taken as a result), and (3) intentionally incorporate clear, consistent feedback between administrators, core team members, and advisors and other end users. Among our sample, the people being asked to use predictive analytics to guide interventions with students were the most concerned about lacking information about how predictive analytics works. Their experiences indicate a clear need for greater care and attention to the practical aspects of how predictive analytics ought to be implemented. Without sufficient attention to understanding end users' needs and providing the information, training, and support necessary to promote buy-in, advisors and other end users are likely to be reluctant to use predictive analytics.

Finally, any efforts to address concerns about the validity or interpretation of predictive analytics should be guided by an awareness of the ethical concerns raised by some of the comments we heard about in this study. An ethical lens should be used to guide the responsible development and use of predictive analytics tools. Attentiveness to the potential for counseling students away from a desired path prematurely, labeling students as failures, and providing legitimacy for racial/ethnic bias should inform the data used in predictive analytics algorithms, the interpretation of outputs, and the guidelines for how the results should be used with students. Making this point forcefully, Ekowo and Palmer (2017, p. 13) ask us to "recognize that predictive-driven interventions can do harm if not used with care."

Predictive analytics has the potential to become a prominent feature of targeted advising in higher education. It is therefore vital that we take heed of the reactions of practitioners who have begun to implement and use predictive analytics tools to ensure that they help college personnel fulfill the goal of advising students wisely and helping them to succeed.

References

Alamuddin, R., Brown, J., & Kurzweil, M. (2016, September 6). *Student data in the digital era: An overview of current practices.* New York, NY: Ithaka S + R. Retrieved from http://www.sr.ithaka.org/publications/student-data-in-the-digital-era/

Beckwith, S. (2016, May 26). Data analytics rising in higher education: A look at four "data czars" and how they're promoting predictive analytics. *University Business.* Retrieved from https://www.universitybusiness.com/article/data-analytics-rising-higher-education

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa

Bryant, G., Seaman, J., Java, N., & Martin, K. (2017). *Driving toward a degree: The evolution of academic advising in higher education. Part 2: The supplier landscape.* Boston, MA: Tyton Partners and Babson Survey Research Group. Retrieved from http://drivetodegree.org/report-archive/driving-toward-degree-evolution-academic-advising-higher-education/

Burke, M., Parnell, A., Wesaw, A., & Kruger, K. (2017). *Predictive analysis of student data: A focus on engagement and behavior*. Washington, DC: National Association of Student Personnel Administrators (NASPA). Retrieved from https://www.naspa.org/images/uploads/main/PREDICTIVE_FULL_4-7-17_DOWNLOAD.pdf

Desouza, K. C., & Smith, K. L. (2016). Predictive analytics: Nudging, shoving, and smacking behaviors in higher education. *EDUCAUSE Review.* Retrieved from https://er.educause.edu/~/media/files/articles/2016/8/erm165w.pdf?la=en

Dimeo, J. (2017, July 19). Data dive. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/digital-learning/article/2017/07/19/georgia-state-improves-student-outcomes-data

Dobo, N. (2017, March 1). Students' worry: Education technology might predict failure before they have a chance to succeed. *The Hechinger Report.* Retrieved from http://hechingerreport.org/students-worry-education-technology-might-predict-failure-chance-succeed/

EAB. (2017, March 8). *Case study compendium: Highlighting diverse institutions that achieved breakthrough results with SSC*. Retrieved from https://www.eab.com/technology/student-success-collaborative/resources/ssc-case-study-compendium-2017

Ekowo, M. (2015, December 14). The gift that keeps on giving: Why predictive analytics are probably on colleges' wish lists this year. *EdCentral*. Retrieved from https://www.newamerica.org/education-policy/edcentral/gift-keeps-giving-predictive-analytics-probably-colleges-wish-lists-year/

Ekowo, M., & Palmer, I. (2016, October). The promise and peril of predictive analytics in higher education: A landscape analysis. *New America*. Retrieved from http://files.eric.ed.gov/fulltext/ED570869.pdf

Ekowo, M., & Palmer, I. (2017, March 6). Predictive analytics in higher education: Five guiding practices for ethical use. *New America*. Retrieved from https://www.newamerica.org/education-policy/policy-papers/predictive-analytics-higher-education/#

Fain, P. (2015, June 2). Phase two of completion. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/news/2015/06/02/sinclair-community-colleges-15-years-completion-projects-pay

Georgia State University. (2014). *GPS advising at Georgia State University*. Retrieved from http://oie.gsu.edu/files/2014/04/Advisement-GPS.pdf

Hoover, E. (2015, March 9). Spotlight on retention: Students can't graduate if they don't return. *Chronicle of Higher Education.* Retrieved from http://www.chronicle.com/article/Spotlight-on-retention-/228173

Israni, E. T. (2017, October 26). When an algorithm helps send you to prison. *New York Times.* Retrieved from https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html

Johnson, J. A. (2017, February 24). *Structural justice in student analytics, or, the silence of the bunnies*. Presented at the 2017 Digital Sociology Mini-Conference, Eastern Sociological Society Annual Meeting, Philadelphia, PA. Retrieved from https://the-other-jeff.com/wp-content/uploads/2017/02/Structural-Justice-in-Learning-Analytics-ESS.pdf

June, A. W. (2017, March 5). How a university fights to keep students' demographics from becoming their destiny. *The Chronicle of Higher Education.* Retrieved from http://www.chronicle.com/article/How-a-University-Fights-to/239396

Kamenetz, A. (2016, October 30). How one university used big data to boost graduation rates. *NPR*. Retrieved from http://www.npr.org/sections/ed/2016/10/30/499200614/how-one-university-used-big-data-to-boost-graduation-rates

LaRiviere, J., McAfee, P., Rao, J., Narayanan, V. K., & Sun, W. (2016, May 25). Where predictive analytics is having the biggest impact. *Harvard Business Review.* Retrieved from https://hbr.org/2016/05/where-predictive-analytics-is-having-the-biggest-impact

McNeal, M. (2016, December 15). Can financial advisors help more students graduate? One university thinks so. *EdSurge*. Retrieved from https://www.edsurge.com/news/2016-12-15-can-financial-advisors-help-more-students-graduate-one-university-thinks-so

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown.

Parikh, R. B., Obermeyer, Z., & Bates, D. W. (2016, April 21). Making predictive analytics a routine part of patient care. *Harvard Business Review.* Retrieved from https://hbr.org/2016/04/making-predictive-analytics-a-routine-part-of-patient-care

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf

Phillips, B. C., & Horowitz, J. E. (2017, April 9). Big data alone won't help students: Educators must know how to make sense of and use the information. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Big-Data-Alone-Won-t-Help/239707?cid=RCPACKAGE

Reed, M. (2017, February 4) Predictions, probabilities, and placebos. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/blogs/confessions-community-college-dean/predictions-probabilities-and-placebos

Renick, T. (2016, January 27). Data fueling scale and change in higher education. *Impatient Optimists*. Retrieved from https://www.impatientoptimists.org/Home/Posts/2016/01/Data-Fueling-Scale-and-Change-in-Higher-Education#.WcaD4cZrzIX

Salisbury, M. (2017, April 9). Big hopes, scant evidence. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com/article/Big-Hopes-Scant-Evidence/239710?cid=RCPACKAGE

Slade, S. (2016, September 6). *Applications of student data in higher education: Issues and ethical considerations*. New York, NY: Ithaka S+R.

Steinhauer, J. (2017, June 7). The telltale data that can identify college students at risk. *New York Times.* Retrieved from https://www.nytimes.com/2017/06/07/education/at-risk-college-students-early-intervention.html

Treaster, J. B. (2017, February 2). Will you graduate? Ask big data. *The New York Times*. Retrieved from https://www.nytimes.com/2017/02/02/education/edlife/will-you-graduate-ask-big-data.html

Warner, J. (2016, September 25). The roots of those not good enough C's. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/blogs/just-visiting/roots-those-not-good-enough-cs

Zinshteyn, M. (2016, November 1). The colleges are watching: With access to predictive analytics and more data than ever before, how can universities avoid invading students' privacy while promoting academic success? *The Atlantic*. Retrieved from https://www.theatlantic.com/education/archive/2016/11/the-colleges-are-watching/506129/