Guidance Counselor Reports of the ASSISTments College Prediction Model (ACPM)

Jaclyn Ocumpaugh,
Ryan S. Baker
Stefan A. Slater
Graduate School of Education
University of Pennsylvania
3700 Walnut St., Phil., PA 19104 USA
+1 (215) 573-2990
{jlocumpaugh,
ryanshaunbaker}@gmail.com

Cristina Heffernan
Neil Heffernan
Department of Computer Science
Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA 01609 USA
{ch, nth}@wpi.edu

Maria O.C.Z. San Pedro
ACT, Inc.
2122 ACT Circle
Iowa City, IA 52245 USA
sweetsp@gmail.com

M. Aaron Hawn
Teachers College, Columbia University
525 W. 120th St.
New York, NY 10027 USA
mah102@tc.columbia.edu

ABSTRACT
Advances in the learning analytics community have created opportunities to deliver early warnings that alert teachers and instructors when a student is at risk of not meeting academic goals [6], [71]. They have also been developed for school district leaders [33] and for academic advisors in higher education [39], but other professionals in the K-12 system, namely guidance counselors, have not been widely served by these systems. In this study, we use college enrollment models created for the ASSISTments learning system [55] to develop reports that target the needs of these professionals, who often work directly with students, but usually not in classroom settings. These reports are designed to facilitate guidance counselors' efforts to help students to set long term academic and career goals. As such, they provide the calculated likelihood that a student will attend college (the ASSISTments College Prediction Model or ACPM), alongside student engagement and learning measures. Using design principles from risk communication research and student feedback theories to inform a co-design process, we developed reports that can inform guidance counselor efforts to support student achievement.

CCS Concepts
Applied computing → Computer-managed instruction

Keywords
Intelligent tutoring systems, stakeholder reports, predictive analytics, guidance counselors, college attendance, student engagement.

Conference'10, Month 1–2, 2010, City, State, Country.

DOI: http://dx.doi.org/10.1145/12345.67890

1. INTRODUCTION
Learning analytics, which has long provided tools for modeling knowledge states (e.g., [15]), has now advanced to the point that real-time engagement indicators (e.g., affective states [48]) and long-term outcomes (e.g., predictions of college attendance [55], [56]) are also becoming common measures. In addition to driving basic research, these models have proven to have a wide range of practical applications. They have been embedded in automated personalization approaches [10], [7], [20] and used to generate reports for both students (e.g., [7], [8], [45], [71]) and education professionals (e.g., [23], [6], [39]). However, as the learning analytics community provides more sophisticated measures, understanding how best to communicate these findings to a wide range of audiences is of increasing importance.

In this paper, we use a co-design process [51] to develop an early warning system for school guidance counselors using data from student interactions with ASSISTments, an intelligent tutor for middle school mathematics [53]. These reports leverage the extensive development of cross-validated student models already available to Learning Analytics researchers who are studying ASSISTments data. Specifically, they use of models of student engagement and learning (knowledge, gaming the system, carelessness, off-task behavior, boredom, confusion, frustration, and engaged concentration). These models, which were further refined to ensure population validity [48], have been used to predict state standardized exams [49], college attendance [55], college major [56], and the selectivity of the college attended [57]. There has been relatively little work, however, to provide data on these types of fine-grained models to school personnel, who could use them in data-driven decision-making. In this paper, we discuss our efforts to use these models to provide guidance counselors, who are responsible for a wide array of educational decisions that impact students' lives, with learning and engagement information that might not otherwise be available in a typical student's file.

Given the large body of research demonstrating that students' trajectories towards college enrollment and success begin years before the college application process [38], there is considerable potential to improve outcomes through an early warning system that targets students in the middle school years. However, research on student feedback suggests that this type of information must be presented carefully [32], [69]. Students need assistance interpreting the feedback (both in terms of interpreting the individual constructs, and in terms of understanding the correlational nature of these models), and they need help setting concrete goals based on this information.

Guidance counselors, who are often over-burdened in their job duties, could benefit from richer data about student engagement with specific domains such as mathematics, which would help them to better support students in preparing for college or for setting alternative career goals. For this reason, in conjunction with four middle school guidance counselors, we have developed two reports based on [55]'s college enrollment prediction model (referred to in this paper as the ASSISTments College Prediction Model *or* ACPM). The first is the *Individual Forecast Report*, which provides each student's likelihood of enrolling in college (as predicted by the ACPM) as well as information about which indicators of learning and engagement (the features used to generate the ACPM) are most heavily contributing to each students' prediction. The second is the *Group Summary of Lowest Performing Factors*, which allows guidance counselors to take a group of students (e.g., all those with a low chance of attending college *or* all those in a particular class) and to identify which learning and engagement factors are most in need of interventions for these students.

2. Background

2.1 ASSISTments

ASSISTments is an intelligent tutoring system designed to assess students' mathematics knowledge while it assists in learning with automated scaffolding and hint messages [53]. ASSISTments is a useful context to conduct this type of research, as it already provides a wide variety of reports on individual students and at the classroom level. Currently, these reports are largely geared towards helping teachers address specific learning objectives (e.g., has a student mastered a specific skill) or towards supporting grading and grade-book management goals (e.g., automatically tallying correctness and assignment completion). However, ASSISTments also has reports in place for parents on homework completion and performance [12]. While ASSISTments has not yet offered reports to guidance counselors, it has the infrastructure in place to do so.

2.2 Guidance Counselors

Guidance counselors provide advice on academic, career, college readiness, and other competencies to students, as part of the school professional community [2]. In the US, these positions were initially created solely to provide support the college application process, but their roles have changed substantially over the years (see reviews in [36], [41], [54]).

Today, 24 of 50 US states mandate that schools provide guidance counselors [3], however guidance counselors often have to support students while dealing with extremely high counselor to student ratios (e.g., 1:800) [2]. Moreover, in schools where they are present, counselors are tasked with a wide range of data-driven jobs, such as (a) helping principals to identify/resolve student needs, (b) advocating for students during meetings that involve future academic or professional plans, (c) providing individual and small/group services that support social development and learning, (d) counseling students with behavior problems, (e) providing individual students with academic program planning, and (f) collaborating with teachers in order to develop effective classroom management strategies [2]. As guidance counselors are increasingly expected to take on more administrative roles [36], they are overburdened with clerical activities and tasks outside their core role, including attendance monitoring, hall monitoring, data entry, and many other support activities for which schools are understaffed [13], [50], [11], [28].

An early warning system that provides learning and engagement data could supplement data from parents and teachers and student self-report data that busy guidance counselors typically rely upon. The use of such a system could significantly improve opportunities to identify the students who are most in need of services, ensure that students are being appropriately challenged, develop programs that address student difficulties in dealing with confusion, frustration, or boredom, identify teachers who are need of support, and otherwise ensure that students are being given adequate opportunities to prepare for college.

3. College Enrollment Prediction Model

Our reports to guidance counselors are built from San Pedro et al.'s (2013) ACPM model that infers college enrollment. The ACPM uses a *discovery with models* approach, where one model is used as a component in another model (see review in [26]).

Specifically, the ACPM was developed by applying models of student engagement and learning to log files of 3,747 students who had used ASSISTments while they were in middle school during the 2004-05 to 2006-07 school years. In this longitudinal study [55], these models were then used to predict which students enrolled in college several years later, using data from the National Student Clearinghouse, which maintains records on all U.S. college students (http://www.studentclearinghouse.org/).

The final model for college attendance was developed using logistic regression and a backwards elimination procedure that removed non-significant features. The resulting model included six weighted features:

$$\begin{aligned}\log(College\ Enrollment) &= .351 - 1.145\ carelessness \\ &+ 1.119\ knowledge \\ &+ .698\ percent\ correct \\ &+ .261\ number\ of\ first\ actions \\ &+ .217\ confusion \\ &+ .169\ boredom\end{aligned}$$

Readers should note that *student knowledge*, the feature weighted most heavily, causes three features to change their direction in this model (namely *carelessness, confusion*, and *boredom*) relative to their direction when considered individually [55]. For example, as student knowledge increases, *carelessness* becomes positively correlated with college attendance even though it is negatively associated with college attendance when considered individually. Under standard 5-fold cross validation (at the student level) this model achieved an A' of 0.686 and a Kappa of 0.247 [55], indicating that it can generalize reasonably well to new students.

4. Co-Design & Design Principles

In order to develop a report that could effectively communicate the college enrollment predictions and the reasons why a prediction is made for specific students to busy guidance counselors, we used a modified co-design process. In traditional co-design, practitioners are included throughout the design process [51]. In this case, we worked with counselors to determine what kind of data would be most useful, but chose to present them with several preliminary design options, rather than to hold design meetings with them where the team started from scratch.

This modification to the typical co-design process was motivated by two primary concerns. First, there is a large literature on the communication of risk which suggests that the simpler designs typically preferred by end users are sometimes inadequate for communicating both the risk involved and the certainty of the model [5]. Second, because this project involved guidance counselors, a group whose schedules are regularly overtaxed, we were reluctant to take more of their time in designing than we absolutely had to. Thus, we sought to utilize design principles from previous research on risk communication and educational feedback to create first designs, presenting potential representations of the predictions to the guidance counselors who participated in the design process with us. In this way, we were able to leverage the benefits of the co-design process while also ensuring that we did not waste our collaborators' time by asking them to reinvent principles that were already well established in the literature.

4.1 Risk Communication Principles

Risk-communication research influenced both the initial and final designs of our reports. In particular, we consulted well-known research on how different forms of data presentation are interpreted by both experts and novices (cf. [4], [29], [61]), since guidance counselors may have varying levels of training in data analysis. Furthermore, it has long been known that even highly trained professionals can interpret information differently when it is presented in different ways or using different scales (e.g., [59]).

Given the potential risks involved when presenting long-term predictions about students, these concerns were given serious consideration during our initial design process. As such, our initial designs drew on several principles that are common in the information design literature (e.g., [35]), including many related to data visualization (to be discussed in greater detail in Section 6). As the design process evolved, other common design principles from the risk communication literature were also incorporated into our designs, many of which were related to the graphic presentation of the material (e.g., RC#1a-e, Table 1, below).

Many of these principles are best understood in context, and will be discussed more thoroughly in the following sections. However, RC2, which deals with labels, deserves further consideration, as does RC4 (provide baseline risks). These principles will be defined in this section in order to facilitate the more thorough discussions to follow.

Table 1. Design Principles from Risk Communication (RC)

Design Principles		Source
RC1	**Visual Characteristics**	
a	Bar graphs encourage comparisons, but are not optimal for proportional data	[29], [35], [64]
b	Tables inhibit interpretations of precision	[35]
c	Pie charts can communicate part-whole relationships clearly, but only if displaying a small number of categories	[61], [64]
d	Keep scales equal and in the same direction	[30], [5], [64]
e	Related constructs should have matching styles/colors (and unrelated constructs should contrast)	[42], [61]
f	Duplicate information	[61]
RC2	**Labeling Characteristics**	
a	Rely on cultural metaphors to reduce working memory demands	[25], [61], [62], [66]
b	Frame labels to encourage fail-soft interventions	[21], [67]
RC3	**Demonstrate interactions**	[35]
RC4	**Provide baseline risks**	[4], [45]
RC5	**Do not exaggerate precision of predictions**	[35]

4.1.1 Cultural Metaphors (RC2a)

Work in the visual representation of information suggests that users can only process a limited amount of new information, leading many to suggest that designers rely on common cultural paradigms, or metaphors, to aid working memory when presenting data [25]. Examples from the literature that were used in this study include traffic light coding schemes (green is good, red is bad, and yellow urges caution; also see [6]) and the common English metaphor of up is good and down is bad.

4.1.2 Framing Labels to Encourage Fail-Soft Interventions (RC2b)

Research on *attribute framing effects* explores the degree to which equivalent information, when presented with either positive or negative reference points, biases people's judgments [38]. For example, [67] found that when people were presented with mathematically identical hypothetical scenarios, they made riskier choices when the odds were framed negatively (33% of people will not die) than they did when the same information was framed positively (33% of people will be saved).

Similarly, framing research has also examined the effects that labels have on the behavior of those in positions of authority, demonstrating that people are more likely to issue punishments to people whose evaluations are framed negatively than they are to the same people when the evaluations are framed positively [21]. Since guidance counselors are part of the authority structure in a school, and since they often face high ratios of students to guidance counselors, where even the most empathetic counselors may be unable to maintain close relationships with every student in their school, this research was especially relevant. Therefore, we worked to ensure that our reports were framed in a manner that would be most likely to result in fail-soft interventions (e.g. [43]) rather than punitive responses.

4.1.3 Provide Baseline Risks (RC4)

Research suggests that when data is presented without context, it is more difficult for people to understand the risk (e.g., [4]). That is, merely telling people that a student is unlikely to attend college or that the student is struggling with a particular engagement measure is not particularly useful unless it is clear how common this issue is.

4.2 Student Feedback Principles

While studies suggests that student feedback is most effective when it comes from a trusted source [73] like a guidance counselor, other research suggests that careful framing of the feedback is also important. Meta-analysis of the research on feedback interventions (FI) has shown that while they can be highly effective, in over one third of cases they actually reduce performance [32]. This is perhaps not surprising since the characteristics of feedback (e.g., amount, frequency, type, and specificity) are known to interact with both student characteristics (e.g., prior knowledge, self-efficacy, motivation, etc.) and task characteristics (e.g., high cognitive load) in determining effectiveness (see review in [72]).

Research on the on the effects of framing feedback messages demonstrates that positively framed feedback is most effective and suggests these effects are enhanced when performance feedback is paired with interventions that help students to produce concrete goals [32]. Experimental research has shown that *positively-framed* feedback results both in higher self-efficacy and in improved performance compared to feedback that only lists problems [69]. Students who receive positively framed feedback are more likely to self-select increasingly challenging tasks [34], possibly because it allows them to set goals, while those who received negative feedback were more likely to show *avoidance behavior,* where students work to minimize opportunities for negative feedback by any means, sometimes including avoiding the academic task altogether [34]. Meanwhile, negatively-framed feedback is thought to add to students' cognitive load, by requiring them to manage their self-concept while performing challenging tasks [68].

Findings such as these have led researchers to advocate for *sustainable feedback principles* [44]. In line with research investigating the development of growth mindsets (e.g., [14], [22], [63]), these researchers argue that feedback should evaluate the task performance rather than the student. They also recommend that evaluations take place immediately after a relevant task, especially when delivering high-stakes predictions.

Table 2 Student Feedback Principles

Design Principles		Source
SF1	Focus evaluations on the task rather than the learner.	Sustainable Feedback Theory (e.g., [44]); Growth Mindset (e.g., [14], [22])
SF2	Frame evaluations positively.	Feedback Intervention Theory (e.g., [27], [32], [68])
SF3	Facilitate the setting of concrete goals.	Sustainable Feedback Theory (e.g., [44]); Growth Mindset (e.g., 14], [22]), Feedback Intervention Theory (e.g., [32])

4.3 Guidance Counselors' Design Priorities

Four guidance counselors participated in the design process, providing feedback on the kind of data that would be useful to them and the ways in which it should be presented. As discussed above, we used a modified co-design process, leveraging both the expertise that could be produced through participatory design with end-users (guidance counselors) and design principles that reflect effective strategies already established in the research literature. This approach allowed us to ensure both that designs were perceived as useful and that the designs lent themselves to the most accurate interpretations possible.

During the co-design process, we explained the ACPM's features (*knowledge, correctness, carelessness, confusion, boredom, number of first actions*) and solicited advice about the kind of information that was most likely to be useful when providing guidance counselor services. While one counselor suggested that she would only want information about students who were on the cusp of not making it to college (those assessed as having a 40-60% chance of attending college, according to this model), others were interested in having information about all students. As one counselor explained, the first thing he would do would be to find the students he knew best, particularly those who were his top performers, in order to better understand the meaning of the data.

As the design process evolved and counselors viewed prototypes, many of their preferences reflected design principles outlined in previous research. These included risk communication principles, such as keeping scales in the same direction (e.g., [35]) and duplicating graphics with alternative means of presentation (e.g., [35], [19]).

Guidance counselor design priorities also reflected research on the framing of interventions. They expressed concerns that echo longstanding admonitions about prematurely labeling a student (e.g., [18], [24], [47], [1]), stressing the importance of framing the model predictions for individual students in ways that reflect the identification of opportunities rather than the creation of static identities of underachievement (e.g., SF1: labeling with reference to performance and behavior on specific tasks or situations rather than labeling the student more generally, as recommended by growth mindset research). Two counselors explicitly said that using negative labels would be detrimental during discussions with students, either because it would disrupt students' ability to focus or because students would "pull away" from someone who was criticizing them (cf., [68]). A third counselor explained that generic coaching (in his example "Oh come on, you can do better!") was ineffective. Positive labels, he said, would assist in setting tangible and consistent goals that could be celebrated with the student upon achievement, in line with *sustainable feedback* theories [44].

Thus, we worked with the guidance counselors to select positively-framed audience-appropriate labels for the measures of learning and engagement that were used to predict college attendance (*knowledge, correctness, carelessness, confusion, boredom, number of first actions*). However, antonyms (often corresponding to the original published label for the construct) were kept in parentheses in order to help disambiguate what each label meant. During this process, it was determined that while the concept of *correctness* (one of the predictor variables in the model) did not provide actionable information to the guidance counselors, the concept of *gaming the system,* (not in the model, but relatively strongly negatively correlated with college attendance) did provide actionable information that was not available in other parts of the model.

Table 3 shows the resulting labels, which were constructed in line with SF1 to focus on performance and behavior in specific tasks or situations, rather than the learner (e.g., *proficiency on tested skills* rather than *a proficient student*, *meticulousness/carelessness* rather than *a meticulous/careless student*, etc.) The resulting report designs are discussed in greater detail in the next section.

Table 3. New Predictor Labels

Original	Re-Labeled
Knowledge	Proficiency on Tested Skills (Low Proficiency)
Carelessness	Meticulousness (Carelessness)
Confusion	Adequate Help Seeking (Confusion)
Boredom	Interest Level (Boredom)
Number of 1st Actions	High Practice (Low Practice)
Gaming the System	Sincere Effort (Gaming the System)

5. Report Designs and Formalized Principles

Guidance counselors who participated in our process wanted reports for two different purposes. Thus we developed two reports: (1) the *Individual Forecast Reports,* which facilitate individual interventions, such as one-on-one meetings to develop personalized goals and (2) the *Group Summary of Lowest Performing Factor,* which facilitate larger group interventions, such working with teachers to identify areas of improvement that an entire class could strive for.

5.1 Individual Forecast Report

Guidance counselors preferred that the ACPM's prediction (the probability that a given student would attend college) be presented alongside factors contributing to that student's prediction. One concern with this approach is that the ACPM is not guaranteed to be causal and a variety of other factors will influence a specific student's college trajectory (e.g., [40], [58], [52]), but the counselors stressed that part of their responsibilities are helping to set goals that improve learning and engagement regardless of a students' desire to attend college. It also raised concerns because of the complexities involved with communicating a logistic regression model to someone who is not familiar with that algorithm (or with advanced data analysis in general).

Longstanding research shows that tables are the best presentation method when individual values (rather than comparisons across subjects) are important (e.g., [29]). Tables also allow data from multiple sources to be presented simultaneously, providing baseline measures that can help contextualize each student's prediction (RC4). However, in order to present our data in accordance with these and other design principles, including those that caution against over-representing model precision (RC5), the data first had to undergo several conversions.

These conversions will be explained in the remainder of section 5.1, while the evaluation of design principles that apply to the Individual Forecast Report will be discussed in section 5.3. However, it is worth summarizing the overall design of this report (shown in Figure 1), which includes nine columns. In addition to (1) each student's name, (2) their individual College Forecast (CF), and (3) each student's Lowest Performing Factor in the ACPM, there are five columns showing normalized performance for the features that comprise the ACPM. As summarized above (in Table 3), these include (4) *Proficiency on Tested Skills* (formerly *Knowledge*), (5) *High Practice* (formerly *Number of 1st Actions*), (6) *Meticulousness* (formerly *Slip* or *Carelessness*), (7) *Interest Levels* (formerly *Boredom*), and (8*) Adequate Help Seeking* (formerly *Confusion*). Finally, the last column shows (9) *Sincere Effort* (formerly *Gaming the System*), which was not included in the ACPM but was individually correlated with college enrollment.

	College Forecast CF	Lowest Performing Category (excludes Sincere Effort, which was not used to calculate CF) LPC	Proficiency on Tested Skills (Low Proficiency) P	High Practice (Low Practice) HP	Meticulousness (Carelessness) M	Interest Levels (Boredom) I	Comprehension (Confusion) C	Sincere Effort* (Gaming the System) *not used to predict CF SE
Alice Bly	80-100%	I	+	+	+	+	+	-
Arthur McBride	60-80%	HP	+	+	+	+	+	+
Flora West	60-80%	M	+	+	-	avg.	avg.	-
Ira Hayes	80-100%	I	+	+	+	+	+	+
Jack Davey Black	60-80%	M	+	+	+	+	+	+
Lamar Houston	60-80%	C	+	+	+	+	+	+
Nettie Moore	80-100%	I	+	+	+	+	+	+
Tim Angel	60-80%	HP	+	+	+	+	+	+
Charlie Patton	40-60%	C	avg.	avg.	+	avg.	avg.	avg.
Dusty Blackcoat	40-60%	P	avg.	avg.	+	+	+	avg.
Frankie Lee	40-60%	P	-	avg.	+	avg.	avg.	-
Hollis Brown	40-60%	P	avg.	avg.	+	avg.	avg.	+
Joe Diamond	40-60%	M	+	avg.	-	+	+	+
Jim Jones	40-60%	P	-	avg.	+	avg.	+	+
Joey DelRey	40-60%	P	avg.	avg.	+	avg.	+	avg.
John Harding	40-60%	M	+	avg.	-	+	+	+
Lenny Bruce	40-60%	HP	avg.	avg.	+	avg.	avg.	avg.
Maggie Farmer	40-60%	M	+	+	-	avg.	avg.	avg.
Maria Washington	40-60%	P	-	avg.	+	avg.	avg.	avg.
Sara J. McMillan	40-60%	C	avg.	avg.	-	+	+	+
Scarlet Pueblo	40-60%	P	avg.	avg.	+	+	+	-
Willie McTell	40-60%	I	avg.	avg.	+	-	avg.	avg.
Duquesne Whistle	0-20%**	M	+	avg.	-	-	-	avg.
Hazel Love	0-20%**	M	-	-	-	-	-	-
Henry Lee	20-40%	M	+	avg.	-	-	-	-
Sally Sue Brown	20-40%	M	+	avg.	-	-	-	-

Figure 1. Individual Forecast Report

5.1.1 *College Forecast (CF)*

The first conversion involved the confidence interval generated by the ACPM. For each student, this value was converted into a percentage and grouped into five ranges (0-20%, 21-40%, 41-60%, 61-80%, or 81-100% chance of attending college), known as the College Forecast (CF). These predictions can be seen in the second column (after each student's name) in Figure 1, where they are also color-coded.

The decision to use these groupings was made in order to minimize comparisons between individual students and to avoid over-representing the precision of the model (e.g., [35]). For example, the difference between a student who is forecasted to have a 63% chance of attending college and a student forecasted to have 65% chance of attending college is likely to be inconsequential, and well-within the model's margin of error.

5.1.2 *Lowest Performing Factor (LPF)*

In the third column of the Individual Forecast Report, each students' the lowest performing factor (LPF) is also identified. This is the factor that is most negatively contributing to each student's probability of attending college. As Figure 1 shows, the labels used in this column matches the column labels for the learning and engagement factors in the following five columns (RC1e).

5.1.3 *Learning and Engagement Factors*

Next, we sought to communicate the degree to which each student's risk of not attending college is increased by each of the learning and engagement factors.

Because the varying scales and coefficients for each factor in the ACPM could make interpretation challenging (RC1d), we calculated what we will refer to as *contribution weightings*—normalized values that reflect the weighting of each feature in the ACPM for a specific student's prediction. Specifically, we took the value of the feature for that student, multiplied it by the weight in the logistic regression equation for that feature, and ran it through a logistic transformation (as was done in the original logistic regression equation). These contribution weightings therefore range from 0 (least contributing to specific student's prediction of college attendance) to +1 (most contributing to specific student's prediction of college attendance).

This process was conducted for the 5 features in [55]'s model (the ACPM). For gaming the system, which was not included in the ACPM, but which was shown to correlate with lower rates of college enrollment [55], we used a mathematically equivalent process, simply computing the value of a single-feature logistic regression model based on gaming the system for this student, also resulting in a -1 to +1 rating for each student.

As with the CF, normalized scores for each of the learning and engagement features were grouped into three ranges. Table 4 explicates the traffic-light color scheme (RC2a) and associated labels (RC1f) used in Figure 1.

Table 4. Representation of Contribution Weightings in the Individual Forecast Report

Contribution Weightings	Color	Label	Interpretation
0 *to* 0.33	red	-	intervention needed
0.33 *to* 0.66	yellow	avg.	intervention could help
0.66 *to* 1	green	+	no intervention needed

This coding scheme matches that of a related construct (the reporting of each student's CF), where data was grouped into five ranges rather than three, supplementing it with comparable colors (RC1e). (Recall that the CF used five categories: green for students with over an 80% chance of attending college, light-green for students in the 60-80% range of attending college, yellow for students in the 40-60% range, orange for students in the 20-40% range, and red for students with under a 20% chance of attending college.)

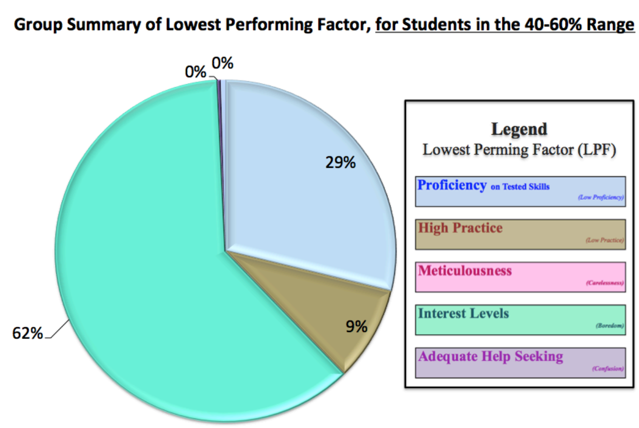
5.2 Group Summary of Lowest Performing Factor

In order to assist guidance counselors in developing interventions for groups of students and/or in providing pedagogical advice to teachers, we developed a second report, the Group Summary of Lowest Performing Factor. This allows guidance counselors who have identified a particular group of interest (e.g., those in the 40-60% prediction range or those in a given classroom) to determine which factor or factors are most in need of interventions for that group.

These reports are designed to provide baseline information on selected groups of students (RC4). In doing so, these reports serve to contextualize reports on individual students, demonstrating how common it is for other students in the school or sub-population to be struggling with a particular predictor variable. As with the individual reports, it is important that these reports encourage interventions which are fail-soft (e.g., unlikely to be harmful if not relevant to every student in the group [43]).

In order to help our guidance counselor to quickly identify the most pressing needs of students in a given range, we used pie charts to show which of the model features were most negatively influencing college predictions, which we called the *Lowest Performing Factor (LPF)*. Figure 2 shows the aggregated information for students in the 40-60% prediction range at one school.

Figure 2. *Lowest Performing Factor (LPF)* distribution for students with a college forecast of 40-60%



5.3 Discussion: Design Principles Applied

With any design, there are trade-offs. End users often prefer simple designs that seem easy to read, even though they may not communicate the data as precisely as more complicated designs [31]. As a result, it is important that co-design processes incorporate opportunities to look carefully at what users do, rather than relying exclusively on what they say they prefer, which often trends towards more aesthetically pleasing, less precise designs. As discussed above, we worked to identify design principals from both the risk communication (RC, as summarized in Table 1) literature and the student feedback (SF, as summarized in Table 2) literature before presenting potential designs to the guidance counselors who collaborated with us. This section discusses how those principles were applied in our reports.

5.3.1 Individual Forecasts

The Individual Forecasts in this study are meant to present information about each student's chances of attending college. As discussed above, our conversations with guidance counselors suggested that providing details about students' learning and engagement could be as important as the College Forecasts (CFs), since this data would help to determine the most appropriate interventions.

Preliminary designs of the *Individual Forecasts* followed advice to use tables rather than graphs (RC1b) to encourage counselors to look up individuals rather compare between students, and also bundled predictions for both the CF and the learning/engagement features (grouping predictions and measures into ranges, rather than providing raw numbers, for ease of interpretation). In addition to facilitating a simpler design (e.g., [60]), this also conformed to research principles cautioning that reports should not exaggerate the level of precision (RC5, [35]).

We also followed research principles related to the use of cultural metaphors in this design (RC2a). Specifically, we used a traffic-light coding scheme (e.g., [6]) where red = low performance, yellow = caution, and green = good performance to highlight differences in student performance on learning and engagement features, as discussed above.

Our final reports also duplicate these graphic (color) representations with alternative means of presentation (RC1f). This was done in several different ways. First, a plus/average/minus labeling system was applied to the learning and engagement measures, duplicating their green/yellow/red coding scheme. In addition to following an important design principle, this also had the benefits increasing the accessibility of the report for individuals with visual impairments and also making the report easier to interpret when printed, since many schools budgets limit their printing options to black/white representations. This same principle was applied to the CF ranges, so that the dark-green/light-green/yellow/orange/red coding scheme was duplicated with the following labels: 0-20%, 20-40%, 40-60%, 60-80%, and 80-100% (achieving RC1e). Finally, we approached the labeling of the Lowest Performing Factor in the same way. While we had initially only used abbreviations of the learning and engagement features to identify the LPF for each student, we found that by using a color coding scheme for this data that was also reflected in corresponding column labels, while not as aesthetically pleasing as a plainer design, ultimately facilitated more accurate interpretations.

As predicted by the literature, guidance counselors both reported positive interpretations of these design choices *and* reflected these positive responses in their ability to accurately interpret hypothetical data while reading sample diagrams and discussing the diagrams' use. Even the color-coding scheme used to match the LPF to column titles, which the guidance counselors initially reported as being a bit distracting, was found to be helpful once they began to use the reports to form interpretations.

Preliminary designs did not follow several other principles, including keeping all scales in the same direction (RC1d), which necessarily means that some variables were not positively framed (SF2). However, our co-design process confirmed the benefits of adhering to both principles. For example, not only did guidance counselors report that they preferred positively-framed variables (SF2) that could facilitate goal setting (SF3), they also found it difficult to interpret negatively-labeled factors. As we worked with them to demonstrate how to interpret interactions between the learning and engagement variables in early designs (RC3), there were repeated challenges in interpreting negatively-labeled features. For example, when "boredom" was used as a column label, counselors would alternate between interpreting the +/green labeling system as meaning "low boredom" (as intended) and "high boredom" (an incorrect interpretation). These interpretation difficulties vanished when designs changed to match previously identified principles in the literature.

Finally, we worked to create learning and engagement feature labels that focused on the evaluation of the performance of the task rather than the learner (SF1). However, some feature labels were still ambiguous. (Notably, the use of the word *attempted* was excluded from potential labels because of it had strong and unintended connotations of failure for the guidance counselors, as in students who *attempted* a problem but were incapable of finishing it.) Moreover, negative affect terms like *confusion* and *boredom* do not have clear-cut antonyms. Thus, in addition to providing an explanation of each variable in the legend for the Individual Forecasts, we also grounded each label with an appropriate antonym, given parenthetically in smaller text in each column. This design choice, which is similar to RC5, enabled us to clarify the meaning of each learning and engagement feature, which should also ultimately support guidance counselors in helping students set concrete goals (SF3).

5.3.2 Group Summary of Lowest Performing Factor

The design of the *Group Summary of Lowest Performing Factor* was, in many ways, simpler than that of the Individual Forecasts. Following work from the risk communication literature that suggests that pie charts are effective for communicating whole-part relationships to lay people (RC1b), we created the ability to summarize data for a given group of students (e.g., those in a single class or those in a particular CF range).

Labels and color-coding schemes for the Lowest Performing Factors reflect those used in the Individual Forecasts (RC1e), allowing guidance counselors to quickly move back and forth between the two reports, and the key duplicates the use of both the positively framed labels and the corresponding antonyms (RC1f). In this way, we are able to ensure accurate interpretations of the learning and engagement indicators reported for each group.

6. Discussion and Conclusions

As learning analytics tools become more powerful, their use in the development of practical tools is becoming more common. Reports for instructors have become routine at all levels, and reports for academic advisors in higher education are beginning to become more standard. However, K-12 counterparts of academic advisors, i.e., school guidance counselors, have yet to have reports designed for their particular needs.

Research shows that job descriptions for guidance counselors have become increasingly more data-driven. However, the distillation of sophisticated modeling and analytics has not reached this audience, a notable gap in the resources available to this audience. Thus, as the learning analytics community continues to grow, this project represents a first step in broadening the audience of student reports from those who are typically targeted (students, teachers, and administrators) to include guidance counselors. Reports specifically designed to assist guidance counselors should be given further consideration, and in particular, their efforts to support student development and teacher pedagogical training will benefit from further support.

Within this article, we discuss two reports designed for guidance counselors in schools that use the ASSISTments mathematics learning platform. Specifically, we provide information on students' college trajectories, using predictive analytics models that can be applied at the end of middle school. Importantly, these reports include data on student learning and engagement measures, which will be beneficial to guidance counselors' efforts, even when they are counseling students who ultimately decide not to pursue college.

We further present our development and design process for these reports, including the principles from risk communication and student feedback research that guided our designs. In general, considerable thought and care should go into the design of reports, as less effective design can lead to unintended and ineffective, or even counter-productive, consequences. We anticipate that these discussions will contribute both to the improvement of the designs discussed in this study as well as to the development of new report systems as this community continues to grow.

The next steps of the research presented here are to move from the pilot work we have already done in partnership with guidance counselors, to scaling the use of these reports. In this way, we can better understand whether their use leads to any benefits to students' outcomes, both within the ASSISTments platform and in their educational pursuits beyond.

7. ACKNOWLEDGMENTS

Our thanks to the guidance counselors who participated in our co-design process and to the NSF (Award #DRL-1031398) for supporting this project.

8. REFERENCES

[1] Abikoff, H., Courtney, M., Pelham W., Koplewicz, H. 1993. Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Ab. Child Psych.*, *21*(5), 519-533.

[2] American School Counselor Association (ASCA) 2016a. https://www.schoolcounselor.org/asca/media/asca/home/RoleStatement.pdf Accessed October 2016.

[3] American School Counselor Association (ASCA) 2016b. https://www.schoolcounselor.org/school-counselors-members/careers-roles Accessed October 2016.

[4] Ancker, J., Kaufman, D. 2007. Rethinking health numeracy: A multidisciplinary literature review. *J. American Medical Informatics Association*, *14*(6), 713-721.

[5] Ancker, J., Senathirajah, Y., Kukafka, R., Starren, J. 2006. Design features of graphs in health risk communication: A systematic review. *J. Am. Med. Infor. Assoc.*, *13*(6), 608-618.

[6] Arnold, K., & Pistilli, M. 2012. Course signals at Purdue: using learning analytics to increase student success. *Proc. 2nd Int'al Conf. Learning Analytics & Knowledge* 267-270.

[7] Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., & Woolf, B. P. 2007. Repairing disengagement with non-invasive interventions. *AIED*, 195-202.

[8] Arroyo, I., Woolf, B., Burelson, W., Muldner, K., Rai, D., Tai, M. 2014. A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *Int'al J. Artificial Intelligence in Ed.*, *24*(4), 387-426.

[9] Baadte, C., Schnotz, W. 2014. Feedback effects on performance, motivation & mood: are they moderated by the learner's self-concept?. *Scandinavian J. Ed. Res.*, *58*(5), 570-591.

[10] Baker, R.S.J.d., Corbett, A., Koedinger, K., Evenson, S., Roll, I., Wagner, A, Naim, M., Raspat, J., Baker, D., Beck, J. 2006. Adapting to When Students Game an Intelligent Tutoring System. *8th Int'al Conf. Intelligent Tutoring Systems*, 392-401.

[11] Barry, B. 1984. A comparison of the perceptions of secondary school counselors, principals & teachers regarding the ideal & actual functions of secondary school guidance counselors in Memphis city schools. PhD Memphis State U.

[12] Broderick, Z., O'Connor C., Mulcahy C., Heffernan, N., & Heffernan C. 2012. Increasing parent engagement in student learning using an intelligent tutoring system. *J. Interactive Learning Research.* 22(4), 523-550.

[13] Burnham, J., & Jackson, C. 2000. School counselor roles: Discrepancies between actual practice and existing models. *Professional School Counseling*, 4(1), 41-49.

[14] Butler R. 1987. Task-involving & ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *J. Ed. Psych.,* 79, 474-482.

[15] Corbett, A., Anderson, J. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling & User-Adapted Interaction,* 4, 253-278.

[16] Corwin, Z., Venegas, K., Oliverez, P., Colyar, J. 2004. School counsel how appropriate guidance affects educational equity. *Urban Education*, *39*(4), 442-457.

[17] Cianci, A., Schaubroeck, J., McGill, G. 2010. Achievement goals, feedback, and task performance. *Human Performance*, *23*(2), 131-154.

[18] Cicourel, A., Kitsuse, J. 1963. The educational decision makers. Indianapolis, IN: Bobbs-Merrill.

[19] Dieckmann, N. F., Slovic, P., Peters, E. 2009. The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. *Risk Analysis*, *29*(10), 1473-1488.

[20] D'Mello, S., Craig, S, Fike, K, Graesser, A. 2009. Responding to learners' cognitive-affective states with supportive & shakeup dialogues. *Human-computer interaction. Ambient, ubiquitous & intelligent interaction.* Springer: Berlin, 595-604.

[21] Dunegan, K. 1996. Fines, frames, and images: Examining formulation effects on punishment decisions. *Organizational Behavior & Human Decision Processes*, *68*(1), 58-67.

[22] Dweck, C. 2008. Mindsets and math/science achievement.

[23] Feng, M., Heffernan, N. 2006. Informing teachers live about student learning: Reporting in the Assistment system. *Tech. Instruction Cognition & Learning*, *3*(1/2), 63.

[24] Foster, G., Ysseldyke, J. 1976. Expectancy and halo effects as a result of artificially induced teacher bias. *Contemporary Educational Psychology*, *1*(1), 37-45.

[25] Gerofsky, S. 2011. Ancestral genres of mathematical graphs. *For the Learning of Mathematics*, *31*(1), 14-19.

[26] Hershkovitz, A., Baker, R.S.J.d., Gobert, J., Wixon, M., Sao Pedro, M. 2013. Discovery with Models: A Case Study on Carelessness in Computer-based Science Inquiry. *American Behavioral Scientist, 57*(10), 1479-1498.

[27] Holmes, J. 2008. Effect of Message Framing on Reactions to Feedback Messages, Moderated by Regulatory Focus. MA Thesis. Virginia Tech.

[28] Hutchinson, R., Barrick, A., Grove, M. 1986. Functions of secondary school counselors in the public schools: Ideal & actual. *The School Counselor*, 34(2), 87-91.

[29] Jarvenpaa, S., Dickson, G. 1988. Graphics & managerial decision making: Research based guidelines. *Communications of the ACM*. 764-774.

[30] Johnson E., Payne J, Bettman J. 1988. Information displays & preference reversals. *Organ. Beh. Human Decis. Process.* 42, 1-21.

[31] Karvonen, K. 2000. The beauty of simplicity. *Proc. ACM Universal Usability*. 85-90.

[32] Kluger, A., DeNisi, A. 1996. The effects of feedback interventions on performance: a historical review, a metaanalysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254.

[33] Knowles, J. 2015. Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. *J. Educational Data Mining*, 7(3), 18-67.

[34] Krenn, B., Würth, S., Hergovich, A. 2013. The impact of feedback on goal setting and task performance. *Swiss J. Psych*.

[35] Kosslyn, S.M. 2006. *Graph design for the eye and mind*. OUP USA.

[36] Lambie, G., Williamson, L. 2004. The challenge to change from guidance counseling to professional school counseling: A historical proposition. *Prof. School Counseling*, 124-131.

[37] Lent, R., Brown, S., Hackett, G. 1994. Toward a unifying social cognitive theory of career and academic interest, choice, & performance. *J. Voc. Behavior*, *45*(1), 79-122.

[38] Levin, I.P., Gaeth, G.J. 1988. How consumers are affected by the framing of attribute information before and after consuming the product. *J. Consumer Res.*, *15*(3), 374-378.

[39] Lonn, S., Krumm, A., Waddington, R., Teasley, S. 2012. Bridging the gap from knowledge to action: Putting analytics in the hands of academic advisors. *2nd Int'al Learning Analytics & Knowledge*, 184-187.

[40] Ma, Y. 2009. Family socioeconomic status, parental involvement, & college major choices—gender, race/ ethnic, & nativity patterns. *Sociological Perspectives*, *52*(2), 211-234.

[41] McKillip, M., Godfrey, K., Rawls, A. 2012. Rules of engagement: Building a college-going culture in an urban school. *Urban Education*.

[42] Misra, R., Mark, J., Khan, S., Kukafka, R. 2010. Using design principles to foster understanding of complex health concepts in consumer informatics tools. *AMIA Annual Symposium Proc.* 492.

[43] Mitchell, T.M. 1993. Machine Learning. Boston, MA: McGraw-Hill.

[44] Molloy, E., Boud, D. 2014. Feedback models for learning, teaching and performance. *Handbook of Research on Ed. Comm. & Technology.* Springer, NY. 413-424.

[45] Muldner, K., Wixon, M., Rai, D., Burleson, W., Woolf, B., Arroyo, I. 2015. Exploring the impact of a learning dashboard on student affect. *LNCSi.* 9112, 307-317.

[46] Natter, H., Berry, D. 2005. Effects of presenting the baseline risk when communicating absolute and relative risk reductions. *Psych, Health & Medicine*, *10*(4), 326-334.

[47] Nisbett, R., Wilson, T. 1977. The halo effect: Evidence for unconscious alteration of judgments. *J. Personality & Social Psychology*, *35*(4), 250.

[48] Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., Heffernan, C. 2014. Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology, 45* (3), 487-501.

[49] Pardos, Z., Baker, R., San Pedro, M.O., Gowda, S., Gowda, S. 2013. Affective states & state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proc. 3rd Int'al Conf. Learning Analytics & Knowledge*, 117-124.

[50] Partin, R. 1993. School counselors' time: Where does it go?. *The School Counselor*, *40*(4), 274-281.

[51] Penuel, W., Roschelle, J., Shechtman, N. 2007. Designing formative assessment software with teachers: An analysis of the co-design process. *Research & Practice in Technology Enhanced Learning*, *2*(01), 51-74.

[52] Perna, L. 2006. Studying college access and choice: A proposed conceptual model. *Higher Education,* 99-157.

[53] Razzaq, L., Patvarczki, J., Almeida, S., Vartak, M., Feng, M., Heffernan, N., Koedinger, K. 2009. The Assistment Builder: Supporting the life cycle of tutoring system content creation. *IEEE Transactions Learning Tech.*, *2*(2), 157-166.

[54] Rosenbaum, J., Miller, S., Krei, M. 1996. Gatekeeping in an era of more open gates: High school counselors' views of their influence on students' college plans. *Am. J. Ed.*, 257-79.

[55] San Pedro, M.O.Z., Baker, R.S.J.d, Bowers, A., Heffernan, N. 2013. Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. *6th Int'al Conf. Educational Data Mining*, 177-184.

[56] San Pedro, M.O.Z., Baker, R., Heffernan, N., Ocumpaugh, J. 2015. Exploring College Major Choice and Middle School Student Behavior, Affect and Learning: What Happens to Students Who Game the System? *5th Int'al Learning Analytics & Knowledge.* 36-40.

[57] San Pedro, M.O.Z., Ocumpaugh, J., Baker, R., Heffernan, N. 2014. Predicting STEM & Non-STEM College Major Enrollment from Middle School Interaction with Mathematics Educational Software. *7th Int'al Conf. Educational Data Mining*, 276-279.

[58] Sandefur, G., Meier, A., Campbell, M. 2006. Family resources, social capital, and college attendance. *Social Sci. Research*, *35*(2), 525-553.

[59] Schwarz, N., Bless, H., Bohner, G., Harlacher, U., Kellen-benz, M. 1991. Response scales as frames of reference: The impact of frequency range on diagnostic judgments. *Applied Cognitive Psychology*, *5*(1), 37-49.

[60] Schwabish, J. 2014. An economist's guide to visualizing data. *Journal of Economic Perspectives*, *28*(1), 209-233.

[61] Shah, P., Hoeffner, J. 2002. Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*(1), 47-69.

[62] Sharp, J. 2011. Semiotics as a theoretical foundation of information design. *Proc. of CONISAR*, 1-5.

[63] Shute, V. 2008. Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189.

[64] Simkin, D., Hastie, R. 1986. An information processing analysis of graph perception. *J. Am. Stat. Assoc.* 82: 454–65.

[65] Sparrow, J.1989. Graphical displays in information systems: some data properties influencing the effectiveness of alternative forms. *Beh. & Information Tech.*, *8*(1), 43-56.

[66] Tversky, B. 2005. Prolegomenon to scientific visualizations. *Visualization in Sci. Ed.* Springer: Netherlands, 29-42.

[67] Tversky, A., Kahneman, D. 1981. The framing of decisions & the psychology of choice. *Environ'al Impact Assessment, Tech. Assessment, & Risk Analysis.* Springer: Berlin.107-29.

[68] Vancouver, J., Tischner, E. 2004. The effect of feedback sign on task performance depends on self-concept discrepancies. *Journal of Applied Psychology*, *89*(6), 1092.

[69] Van de Ridder, J., Peters, C., Stokking, K., d Ru, J, tenCate, O. 2015. Framing of feedback impacts student's satisfaction, self-efficacy & performance. *Adv. in Health Sci. Ed.*, *20*(3), 803-816.

[70] Vessey, I. 1991. Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature. *Dec. Sci.,* 22:219-40.

[71] Walonoski, J., Heffernan, N. 2006. Prevention of off-task gaming behavior in intelligent tutoring systems. *Intelligent Tutoring Systems.* Springer: Berlin. 722-724.

[72] Wielenga-Meijer, E., Taris, T., Kompier, M., Wigholdus, D. 2010. From task characteristics to learning: A systematic review. *Scandinavian J. Psych. 51(5), 363-375.*

[73] Wiliam, D. 2010. The role of formative assessment in effective learning environments. *The nature of learning: Using research to inspire practice*, 135-155.