# Taking Student Data for Granted? A Multi-Stakeholder Analysis of a Learning Analytics System

Kaiwen Sun, Abraham H. Mhaidli, Sonakshi Watel,
Christopher Brooks and Florian Schaub
School of Information
University of Michigan
Ann Arbor, MI, USA
{kwsun, mhaidli, sonakshi, brooksch, fschaub}@umich.edu

## ABSTRACT

Early Warning Systems (EWS) are a critical topic of interest for educational data mining and learning analytics. EWS use large-scale educational data to generate reports or interventions to understand and guide learning behaviors. As EWS are becoming a more integrated part of the higher education experience, it is important to understand different stakeholders' perceptions, attitudes, and expectations toward the ethics and impact of the access, use, and analysis of learners' data in such applications. In this work, we take a multi-stakeholder approach to understand perceptions on Student Explorer, an EWS deployed at the University of Michigan. We conducted 32 semi-structured interviews with individuals from three stakeholder groups. Our findings indicate both consistent and inconsistent attitudes and perceptions among stakeholders toward Student Explorer and the use of learners data.

## Keywords

Educational Data Mining, Learning Analytics, Early Warning Systems, Higher Education

## 1. INTRODUCTION

Educational data mining is concerned with the analysis of large-scale educational data for various purposes, including the generation of reports or interventions to understand and guide learning behaviors, such as Early Warning Systems to identify under-performing students [3]. Educational data mining has been incorporated with learning analytics systems that are designed for different stakeholders, each of whom may have different roles within the overall educational system – such as students or academic advisors. As these systems become a more integrated part of the higher education experience, associated stakeholders' perceptions, attitudes, and expectations towards the access, use, and analysis of learners' data involved in the educational data mining and learning analytics process are not well studied.

To understand different stakeholders perceptions of learning analytics systems and the use of learners data, we conducted 32 semi-structured in-depth interviews with three groups of stakeholders (developers, advisors, and students) of a specific early warning system in operation at the University of Michigan called Student Explorer [6]. Our findings indicate that different stakeholders have various interpretations of what data accuracy means due to different experiences with the systems (human and technical) that generate data; additionally, there is a widespread concern that EWS data is not completely accurate. Furthermore, all stakeholders think that the information displayed in the early warning system is an incomplete representation of a learners' overall academic performance.

Additionally, given each stakeholders unique position in the learning process, there are varying degrees of trust towards the data and each other: trusting whether the data is useful, whether the data will be used ethically by other stakeholders, whether stakeholders can be trusted with access to the data, and so on. We also identify discrepancies among the stakeholders regarding the awareness of the existence and use of the data, indicating a lack of communication among the relevant parties. In addition, we identify tensions among the stakeholders regarding agency and control over the data, with different stakeholders having different expectations over who should be able to control and access what information.

These findings show the need for better involvement, integration, and engagement of all stakeholders in the design of learning analytics systems, specifically early warning systems which rely on the creation of predictive models.

## 2. RELATED WORK

Higher education institutions are increasingly making use of analytics and educational data mining to further goals [2]. By looking into large data sets to draw relevant information, institutions can better understand the learning process [12], developing new and better curricula [4], and increase student success [1]. One prominent use of learning analytics is the creation of Early Warning Systems (EWS). These systems leverage predictive models to identify 'at-risk' students who might be in danger of failing a class or underperforming academically; this allows for timely intervention by the institution or instructor, enabling students to succeed [3, 8].

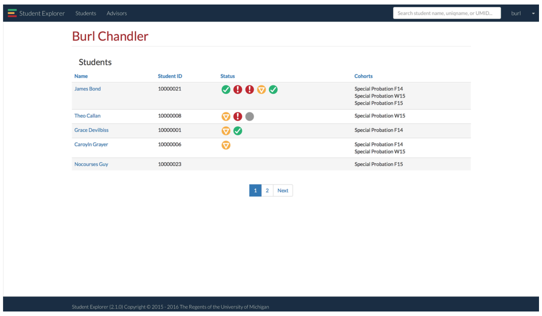


**Figure 1: "Student roster" view after advisor log in Student Explorer.**

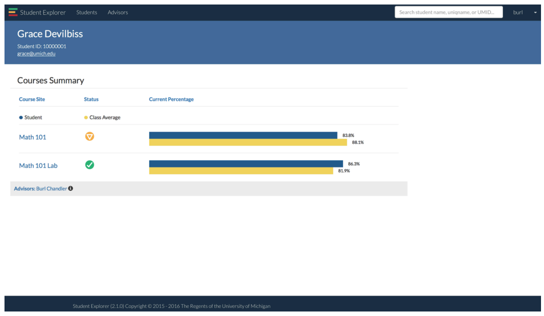


**Figure 2: Student's performance (blue bar) comparing to the class average (yellow bar).**

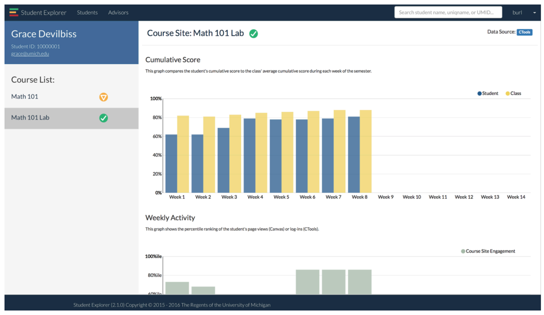


**Figure 3: Student's cumulative score (blue bar) to the class average cumulative score during each week of the semester.**

A big part of creating these EWS, and educational data mining in general, is the collection and processing of vast amounts of data of students, which can include grades, attendance, honors and awards received, summer and post-graduation plans, and time at which certain online assignments are opened by students [5, 9, 2]. This data needs to be collected, aggregated, stored, and processed – which raises significant privacy concerns and questions of who can access the data and how it may be used [9].

To address these concerns, Prinsloo & Slate [10] propose ethics of care for learning analytics, aiming to involve individuals and groups during their personal data's collection process, informing them how their data is used, and granting them access to that data. Another recurring theme is maintaining transparency. This means that students should know how, why, with whom, and for what reasons their data is collected, used or shared [11].

Our research contributes qualitative findings on key stakeholder attitudes and perceptions regarding an EWS. Specifically, we provide insights on the perspective of three groups of stakeholders (developers, advisors, and students) by examining their attitudes, perceptions, and expectations regarding key issues of EWS, such as data collection, data access, data use, and consent.

## 3. STUDENT EXPLORER

Student Explorer [6], is an early warning system deployed at the University of Michigan that leverages Learning Management System (LMS) data from Canvas (by Instructure), a LMS widely used on campus. Student Explorer aims to assist academic advisors in identifying students at risk of academic jeopardy in order to facilitate outreach to these students [7].

When advisors log in Student Explorer directly, the first page they see is the "student roster" view showing a list of students they are assigned to advise, the students' name, student ID, and the students' performance status summary (see, Figure 1). The red, yellow and green symbols represent whether the students' performance is below, on par, or above the class average, respectively. These colored symbols are calculated based on a data extraction process from Canvas.

On clicking a certain student's name, advisors can see all of his or her class performances in comparison to the course average. The colored symbol status (green check-mark, yellow triangle, and red exclamation point) should reflect the comparison result accordingly (see, Figure 2). Clicking on a course opens the course site page, showing the student's weekly cumulative performance across the entire semester against the course average (see Figure 3).

Regarding the relevant stakeholders of Student Explorer, we identify three key groups: the *designers and developers* who built the system, *academic advisors* who are the primary users of the system, and *students* whose data is used in the system. Due to the indirect role they play in the context of SE, we did not consider instructors direct stakeholders.

## 4. STUDY DESIGN

We conducted 32 in-person semi-structured interviews between October 2017 and February 2018 with three groups of key stakeholders of Student Explorer: developers, advisors, and students. All interviews were audio recorded and lasted 35-45 minutes on average. Student participants were compensated $10, advisor and developer participants were not compensated. Our study was approved by our institution's IRB.

### 4.1 Recruitment

Developers were recruited via snowball sampling. Advisors were recruited via mailing lists: we specifically targeted advisors who were working with Student Explorer. Students were also recruited via mailing lists in different deparmtents and were asked to complete a pre-screening survey. We recruited academic advisors and students from multiple colleges and departments across the university. Particularly for student participants, we aimed to balance gender, school year, and majors. In recruitment, we did not mention that the focus of the study was the use of student data to reduce self-selection bias.

Similarly, we began interviews by asking participants about their role, familiarity and use of Student Explorer, including data usage questions in the process. For student participants, we started the interviews by asking about their academic advising experiences, then introduced them to Student Explorer to learn about their perspectives.

## 4.2 Participant Demographics

We interviewed 4 developers, all were male. Of the 8 advisors we interviewed (4 female and 4 male), 4 were in the Literature Sciences and Art Department (with 1 advisor in the Literature Science and Arts Honor Program), 1 in Information, 1 in Engineering, 1 in Comprehensive studies, and 1 in business. Student demographic information is summarized in Table 1.

**Table 1: Student Participant Demographics.**

| ID | Gender | School Year | Department |
|---|---|---|---|
| S1 | M | Senior | History and Communication |
| S2 | F | Sophomore | History in Philosophy Political Science and Econ |
| S3 | M | Freshman | Undecleared |
| S4 | F | Junior | Biopsychology Cognition and Neuroscience |
| S5 | M | Freshman | Undecleared |
| S6 | M | Freshman | Engineering |
| S7 | M | Freshman | Engineering |
| S8 | F | Senior | Environmental Engineering |
| S9 | F | Sophomore | Civil and Environmental Engineering |
| S10 | M | Sophomore | Mechanical Engineering |
| S11 | F | Junior | Information Science |
| S12 | M | Senior | Information Science |
| S13 | F | Senior | Business and Information Science |
| S14 | F | Senior | Information Science |
| S15 | M | Junior | Information Science |
| S16 | M | Junior | Business and Biology |
| S17 | F | Senior | Business and Music |
| S18 | F | Sophomore | Business and Public Health |
| S19 | M | Sophomore | Business and Germany |
| S20 | M | Junior | Information Science |

## 4.3 Analysis

Qualitative data analysis was conducted by three research team members, who conducted thematic coding on the transcribed interviews, followed by affinity diagramming. The coders worked together to identify preliminary themes and establish codebooks (a separate codebook for each group of interviews). The codebooks were collaboratively and iteratively refined through independent coding until sufficiently high inter-rater reliability was reached (Fleiss' $\kappa > .75$). One researcher then coded the remaining interviews and recoded the previous interviews using the final iteration of the codebook.

# 5. FINDINGS

We discuss key findings in five categories: (1) awareness and transparency of student data use in SE, (2) accuracy of the data, (3) holistic and incomplete nature of the data, (4) trust issues between (a) stakeholders and the data and (b) between stakeholders, and (5) agency and control of the data.

To discuss the findings, we assign ID's to the participants. Developers are identified as D1 through D4; advisors as A1 through A8; and students as S1 through S8.

## 5.1 Awareness and Transparency of the Data

SE uses primarily student's data, but are the students aware that their data is used on the learning analytics platform? We found that most of the non-student participants are unsure of students' awareness about their data used in SE. None of the students had ever seen, heard of, or been informed, of SE or how their data is being used by the platform.

Some developers stated that since SE is not designed for the students, students probably are not aware of its existence, and students are not informed their data is used in SE. D1 explained: *"I don't think that [SE] is sort of a widely known system in a lot of ways. [...] Part of that is that it doesn't have a student facing interface. And so ultimately out of sight out of mind [for the students]"*. Other developers shared that some students might be aware of SE because they could have learned it during the advising sessions. D2 said: *"I think when [students] go to advisor, they sometimes ask the advisor to show them [SE], because they are curious to know how they're doing with respect to class average [...] So the ones who come to advisor they know [about SE]."*

For advisors, some are mainly focused on their job responsibilities and are unsure to what extent students are aware of SE: *"Students operate in a particular digital universe, and staff operates in their own [...] sometimes we do mostly focus on my work, and how I used Student Explorer, do students use it, I don't know. I just use it for what I use it for"*, A8 explained. Some advisors don't think the students are aware of SE based on the students' reaction seeing SE. A4 said: *"[Students] always ask, is there a way I can see this, and how do you get this data? Can you see my grades?"*

None of the student participants had heard of SE before the interview, nor been informed that their data is used in SE. S4 states: *"No I don't think I've been informed with Student Explorer like using my data. [...] I feel like being open with students is probably much more helpful than just like using data like secretively."* Most of the student participants think they should be informed when SE is using their data via a transparent process since it is their data that is being used, and so, they should know how it is being used. *"I should be informed and I should be asked for permission, its not like an automatic process, [...] we should have, [...] a formal process where the university asks you to sign if you are willing to share your data"* said S11; *"Student should be aware of this [SE], because otherwise [...] it's just like this really weird thing where [...] your data is out there somewhere and someone who is not you has access to it"*, S2 added.

Despite not knowing students' awareness of SE, some advisors said it would be fair to students to be informed of how their data is used. A5 stated: *"I don't think it's a problem for students to know [their data is used in SE] like I said, if I have a student who's on probation or who you know has some significant issues and isn't doing well in all their classes like I will pull it up and you know talk to them about it."* However, none of the advisor participants thought it was their responsibility to inform the students; instead, they thought other platforms or stakeholders could inform students. For example, A2 said: *"If they [students] participate in Canvas I could imagine at the Canvas level [...] make you aware of your data from your classes is available to [...] advisors"*; A4 shared: *"You know maybe in the first day class instructors could talk about this [SE] as just another data point and a vehicle to help students stay on track."* In contrast, some students considered it to be the advisors' responsibility to inform them that SE is using students' data. A9 shared: *"Maybe advisors could tell students that they have this information, maybe just like the first time that they meet with them, or like have the freshman advisors tell these students."*

We identify discrepancies among student and advisor participants regarding the awareness and transparency of the existence and use of the data in SE. Each group of stakeholder approach this issue from their own perspective and there's a lack of agreement and discourse among the relevant parties to address whose responsibilities it is to involve and inform the students in the learning analytics process.

## 5.2 Accuracy of the Data

Different groups of stakeholders have different experiences with the data in SE; as such, their perspectives vary when it comes to the accuracy of the data. The developers approach the accuracy of data from an algorithmic perspective; the advisor participants' understanding about the accuracy of the data is related to how student data is displayed through icons and graphs; as for students, since it is the first time they have heard or seen SE, their knowledge on the accuracy of the data are more related with SE's data source, Canvas, and less so with SE itself.

We identify from the stakeholders that there are several factors that contribute to the concept "accuracy" and they have interactive relationships with each other: (1) whether SE's calculation accurately captures all of a students' performance, (2) whether the data source of SE, the learning management system Canvas, has complete and accurate data input from the instructional team of each course, and (3) whether SE's visual language (icons, graphs, and charts etc) clearly and accurately reflects students' performance without additional clarification or explanation.

For the first factor, since different stakeholders have different experiences and knowledge of SE, and their touch-points with SE vary due to their unique roles, so their understanding of the accuracy of data can overlap with any one or multiple factors mentioned above. For developers, they created SE and so have the most comprehensive understanding of how the inner algorithms work. Some of them shared that the design of the algorithm is not perfect, since it cannot completely capture and calculate all the data situations. D1 mentioned: *"Student Explorer is ultimately predicated on algorithms and those algorithms do their best to cover the range of cases. But ultimately like any sort of algorithm that's pointed at a variable data set, it probably uses more to the 80:20 kind process [ 80/20 rule that 80 percent of the data situation can be captured easily, while the rest 20 percent of the data situation are more complex to catch]."* Student participant S7, who has some academic knowledge of computer algorithms, also understands the incomplete nature of algorithm in general, thus S7 speculated the algorithm in SE is difficult to cover all the student cases: *"I mean obviously the algorithm [in SE] has to set one up in some way, but in real life you can't really like you have to evaluate each individual case."* In this case, when the algorithm fails to identify some complicated data situations, developers acknowledge the inaccuracy brings limitation to SE's expansion to more stakeholders, some advisor and student participants are concerned that using the inaccurate data to measure students' performance can be misleading and unfair.

The fact that the algorithm in SE is unable to capture all the data situations is closely related with the second factor: whether the data source of SE, the learning management system Canvas, has complete and accurate data input from the instructional team of each course; if not, the algorithm mainly follows certain criteria to extract the data rather than actively calculating it according to the developers. As the data source for SE, Canvas has several issues providing the most accurate data for SE. First, not every single class on campus uses Canvas as the learning platform as S9 stated: *"It's not accurate when [...] if [the instructors] don't use [Canvas]."* Some instructors choose to use other systems or methods as a part of the learning experience in the class, take an example from S5: *"So like in my Portuguese class, we have other online tools [...] which account for like 20 percent of my grade."* For those classes that don't use Canvas or only use part of it, SE doesn't convey from the interface to the advisors that the students' data is incomplete.

Regarding the second factor, the data source, even though many classes use Canvas, instructors don't always know how to properly use the system: inconsistent timing to update student grade, label assignments under different categories freely, or don't update any grade information at all etc. All these factors can affect the data accuracy in SE. A6 said: *"So for the people [instructional team] who are actually inputting the data so making sure that they are aware of how it's being used."* Instructors construct the class with various assessment components, some of them such as participation or group evaluation are not reflected in SE. For instance, S18 said: *"For [business school] class for example like participation is a huge chunk of your grade like it's as much as homework [...] like [the instructional team] don't even put it online at all."* Also, sometimes the instructors give students the options to choose which assignments to do, so students have the flexibility to skip certain assignments, S14 said: *"For classes that are like a bottom-up structure where students can do you know assignments that they're choosing throughout the term, it can be misleading [if the advisor see the student misses a homework without knowing why the student skips it], it is probably not super helpful in that case."* For some instructors, they don't grade or update the assignments scores on time leading to incomplete information on SE as A1 stated: *" Sometimes you will just have the discussion portion of a class put it but not the lecture [...] their exam grades are missing so it is hard to get a sense of where the student is at."* For some assignments, they are feedback oriented as S12 mentioned: *"so some course [...] which only have like one giant paper, so you have like nothing [no score] all the way towards the end"*, which can't be translated into a score reflecting on SE. When it comes to classes with gameful learning elements that allow more personalized learning experience to engage with the students, D1 said: *" the system completely falls over when you start to think about gameful courses."* Such inconsistent Canvas usage behaviors by instructors accompanied with a static algorithm can result in empty or inaccurate data in SE.

For the third factor, SE's visual language, which is the critical communication method for the advisors to rely on to understand students' performance. For some advisor participants, they use the colored symbols (green check-mark, yellow triangle, and red exclamation point) to identify students' performance and make corresponding actions to intervene and guide students' behavior. For the student participants, they establish perception and understanding of SE through the screenshots shown in Figure 1 to Figure 3, so the visual language is important for them to form their own interpretation of the interface. Most of the advisor and student participants can understand the visual language of the colored symbols meaning green is good, red is bad. However, advisors are unsure about what criteria determine when and why the colored symbols change. A2 said: *"I noticed that the definitions of when something becomes yellow and red don't seem to be that consistent."* Also, advisors don't know the severity indicated by the yellow and red symbols, as A8 shared: *"Generally it's like I know it [the yellow symbol] is bad but what does that mean. Or how are you determining. So like a key or a legend might be useful like ' this is how this [the colored symbol] is working."* Student participant S4 also commented: *" The yellow icon is kind of like iffy because I'm not sure whether it's like telling me that the student is around the average or if the student is doing slightly worse, or like overtime or the student is doing better, even though he or she is like still below the class average."* Since there's no key explaining how the colored symbols are determined, or any instruction to guide advisors how to properly interpret the symbols, some advisors don't think the data can accurately inform them how the students are really doing, for example, A6 said: *" I am cautiously using the information. And usually using it as a way for the student to further expand on what's going on"*, and they need further explanation from the students about the data rather than purely relying on the colored symbols. In comparison, the developers are less aware of the inaccuracy caused by SE's visual language, they believe the colored symbol are good indicators to help advisors identify student status and make moves. D3 shared: *" So if they can get a summary view [by looking at the colored symbols] of what's going on it [colored symbols] gives them a real good indication of how the student is doing and then they can proceed with the student."*

Overall, despite of the various interpretation of what data accuracy means, all the stakeholders think the data in SE is not completely accurate.

## 5.3 Holistic and Incomplete Nature of the Data

SE strives to provide a holistic view of the students' class performance summary in an easy to understand way, and it uses student's individual assignment scores and aggregated average scores in the class to determine how the student is doing. In this case, advisors can use such data to identify different students' performances. However, when the data in SE is not completely accurate due to various reasons stated in the previous section, almost all of the stakeholders acknowledge that the data in SE is an incomplete representation of students' overall academic performance.

A6 said: *" The student could have a warning icon [red exclamation symbol] but not all the grades are in the system [SE]"*; S16 gave an example: *" I would think incomplete [...] for example one of my like bio chem classes, if you get a 50 percent on the problem sets, you'll get a 100 percent [...] but that's not really reflected in the grade until the very end of the semester so. I think that it might not be a full complete view of the entire semester based on my own experience"*, in which case the missing data leads to the incomplete picture of the students' performance.

Another perspective shared most of the student participants is that to truly evaluate students' performance, there are so many factors that are less quantifiable than a single data point in SE such as S1 stated: *" How [the students are] contributing to discussion which [...] can't be graded, just like [...] written comments. I mean that's not transferable to numerical on entry."* S17 shared the limitation of quantified data: *" like I'm hoping an advisor can see it's not just your actual GPA like number, it is also your ability to perform[...] I definitely don't see this being represented in here [SE]."* Some student participants mentioned SE might be more suitable, provide a relatively more complete representation of students' performance in certain classes with larger enrollment and more regular automated homework so the scores are more easily and timely updated: *"I feel like [StudentExplorer] would be accurate for larger classes [...] that it's all automated [to calculate assignment scores."* So the effectiveness of the data in SE can be contextually based on how well the class is designed to be evaluated by scores. On the other hand, some student participants stressed the difference between grade and learning: *"Nothing is a complete representation [...] test scores don't necessarily equate to learning you know, and grades don't necessarily equate to learning"*, S7 insisted.

The purpose of SE is using quantified data to measure students' performance as a way to measure learning. All the stakeholders recognize the limitation of the data in SE, which is an incomplete representation of the students' overall academic performance.

Besides grades, students expressed that their classroom engagement, actual ability to perform, career interest, personal goals, soft skill evaluation, instructors' written evaluation, impacts from students personal life, etc, should all be considered to evaluate how the students are learning and performing overall.

## 5.4 Trust Issues

While all stakeholders see data inaccuracy as an issue, their level of trust in SE's data varies. Such variations are due to their respective role in the learning analytics process and their understanding of how the data is used.

### *5.4.1 Varying levels of trust in data*

The developer participants are confident that the data in SE is helpful for assisting advisors in identifying students who are at risk. For instance, D4 stated: *"So that early intervention [...] we saw a lot of improvement there. We heard* *from advisors that they were able to get ahead of these problems much more quickly than they were before."* Developers further saw value for the university to use SE data to understand student performance trends in a broader context: *"A system that basically can track over time when students basically enter different states [...] being able to observe those patterns those trends across multiple courses [...] can give [...] the university lands on the way that courses are constructed [...] measuring how the student experience changes within a course [...] it's certainly been a huge value add to the learning analytics community"* (D1).

However, the advisors did not fully trust the data in SE, but they find their own way to use SE despite the inaccuracy. For example, A2 said: *"When I had a student who tells me all this [data in SE] doesn't match my grade. [...] So I've learned to take that with a grain of salt."* A6 expanded: *"An issue is that it's not completely accurate. So it doesn't have my full trust. So it's like I am cautiously using the information. And usually using it as a way for the student to further expand on what's going on."* Once the advisors find out from the students that the data is not accurate, they become more aware and careful toward SE data when using it. Surprisingly, some of the advisor participants even take advantage of the disadvantage (data inaccuracy) and shifted their mindset as A7 also shared: *"It not being 100 percent reliable I think is ok because it gets me looking a little bit deeper and double checking myself. But if somebody was to just blindly trust it I don't think that would work out well."*

Since the students only have access to Canvas but not SE, they primarily discussed to what extent they trust data in Canvas. A few students displayed trust in Canvas because they use it all the time and can immediately take actions if they see something wrong. S10 said: *"I think students will definitely trust Canvas more than the third party [...] First of all is what we're used to. And it's also if you see the data sources from canvas then, and [SE] is incorrect. Or you see something very strange here [in SE] I guess there's no way of knowing it's incorrect until you compare it with Canvas. But I would definitely trust Canvas."* The student worries if the data in SE is incorrect, that the advisors won't be able to know, unless the student and the advisor communicate with each other, because they can't access Canvas directly.

### 5.4.2 Student trust in advisors and the instiuttion

In addition to the different stakeholders' perspective of trust in the data, participants also talked about trust relationships and interactions among the stakeholders.

Most of the student participants generally trust the advisors to have access to the students' personal and academic data as they understand the advisors will use the information to provide better advice. S16 shared: *"It's just very useful to [advisors] to get a sense of where I am, like academically standing. And just I guess [having advisors access student information] helps them to get a little more prepared for what I come in."* Besides the advisors, there's an invisible "them" that also plays a part in the trust issues according to the students, namely "the university." Some of the students know their data is used by the university for various purposes, and they have mixed feelings about it: some hold the university accountable to use students' data properly for good: *"I think I trust that this are only being used for like supporting students and like supplying data to the university. I mean that the university already has all of this data anyway"*, said S8; while others think they don't have other choices instead of trusting the university: *"I think part of it is trust. Part of it is kind of resignation because it's like I kind of regardless of whether or not kind of I want to trust them [the university] or not. Like they had my information and they can kind of do with it whatever they want. [...] I could stay distrusting but it just gives me personally more peace of mind to say like well you know might as well trust them, they seem like they know what they're doing."*

Above all, trust issues exist in the connected relationships between stakeholders, data and the university. We identify interesting dynamic among the stakeholders and data due to the unique role each group of stakeholders play, their different personal experiences, and the lack of communication.

## 5.5 Agency and Control of the Data

As the educational data supplier, regardless of the learning analytics system's purpose, students play a crucial role in the higher educational system. However, how much do students get to have a say in the usage of their data in SE? What do stakeholders think of students' consent, rights, and options when their data is used in SE? We discover tensions among all stakeholder groups regarding these questions.

Some developer and advisor participants think student consent involves a higher level of discussion in the university. These developer participants present themselves as good stewards of the learners' data but consider questions of consent and transparency beyond their roles and responsibility, D3 said: *"I think that [student consent] is probably a higher level discussion at the University. It is probably not something that we, that Student Explorer would need to address directly because the data itself is coming from Canvas gradebooks and other places that are just the very basic business of the University"*, D4 also shared: *"This is part of a larger conversation about a student's ability to kind of opt in or out of the types of things that are being tracked about them at the university."* For the advisors,they acknowledge the students' right to be informed on how their data is used, but also did not uniformly think that students should be able to opt-out of SE, because they consider that their special role as advisors requires a deeper level of access to student information to provide proper help and guidance. So the advisors consider it critical to have access to SE regardless of the students' preference as A5 commented: *"I think students do have a right to know what is shared. But at the same time because it's for their own benefit that they may not know all the resources that are available to them, then it would be it would be my preference that we still have access to this no matter what a student says."* A8 also shared a similar concern: *"With the freedom of information, I guess in some way [students] could say, no I don't want you to use the [student] data but that kind of closes off the tools that we use. Put it this way we need the data to do their job. But I can understand students wanting or keeping tight reins on information. Basically this is their academic information. [...] I think just because of an individual rights standpoint, if they wanted to, they should be able to turn it off, but we need it. So I can kind of see both sides."*

Most of the student participants believe they should have a say in how their data is used such as having an opt-out option for SE, being asked for permission if system such as SE wants to use their data, know when the data is used by having email notifications etc. S1 mentioned: *"This is a student's education and they should be [...] liable they should be in charge of whatever relevant information regarding academic career holds."* S18 followed: *"Another option would be like I would have permission for this[the access of SE], and I don't give permission to like administration to have my information."* Some student participants have an unsure or no opinion toward student consent because they think the university is using students' data regardless. S14 stated: *" I don't know. I feel like as a student there is a lot of like my data the school uses anyway [...] I don't think that I would feel that uncomfortable with something like this existing online just because I feel like they are there already probably do something like this."* The rest of the student participants think individual student consent is less important comparing to maintain the integrity of the data. *"I would want to say no [student should not have a say in how their data is used] just because I think that if it's university policy to kind of use this for advising, it would make sense to have as much information as you possibly can, because then the algorithm can learn better"* said S19; also these students think it's too much work for both the university and the students if asking every single student for data usage permission as S7 mentioned: *"I mean really. no, I feel like it would it would be such a difficult issue to deal with, you have to just to give a waiver form for every student or something, or like it or implement an option to not allow their data to be used for this. I think that that would be too big of a hassle and I don't think it would be really worth it."*

Although SE is a platform designed for advisors not the students, most of the student participants and a few advisor participants advocate that students should have access to SE. S14 shared: *"It would be nice for students to see their progress against the class [in SE] [...] to see these trends."*. From A7 we learned: *"I think it would be good for most students [if they can access SE]. I think it would be a conversation between them and advisor and instructor about like you know how to use it, what to take away from it"*, indicating that the advisors know it's better to provide students guidance to interpret the data rather than blindly giving them access. This consideration is one of the reasons some developer and advisor participants hesitated to open up SE to the students. D1 explained: *"There were sufficient sort of risks or things that we weren't fully confident on in terms of their impact on students that we didn't make it available to them to have that view of themselves [...] [for example] telling a student that they're in trouble [if a student sees red symbol in SE] without giving them any help or that sort of context or advice around what they should do. It was something that we were cognizant of"*; A6 also shared the concern: *"Because until its [SE's] accuracy is more solid, then I think it should be limited [to student access] because of how it could be easily misinterpreted."*

In terms of students' attitude toward advisors' access to SE and being able to see students' data, most of the student participants understand the advisors are here to provide help. However, a few of the students also worry that giving the advisors access to such detailed data (e.g. students' weekly assignment activities) listed in SE can have negative impacts. To be more specific, depending on the students' relationship with an advisor, or how they view the role that academic advising plays in their life, their trust level on the advisors' access to the data on SE can be affected, S11 commented: *"The advisor looks at your grade, uh, sometimes they would have some presumptions for you because of your grades, its not like they want to do that, sometimes you [advisors] really cant help it, and thats kind of bad"*; S7 also mentioned: *"[Students] just might be uncomfortable in general with people other than themselves seeing their grades even their advisor."* Also, some student participants don't feel comfortable knowing some developers can access SE. S7 said: *"I think it's kind of strange that the developers have access"*, they think there should be more access controls such as S10 said: *"If they anonymize the names that would make even more sense to me."* S15 suggested: *"If you sign up like say [...] I give you permission to use this [student data] you know anonymously [...] on other platforms."*

# 6. DISCUSSION

Before discussing the insights from our findings we shortly address potential limitations of our research.

## 6.1 Limitations

Most of our student interviewees reported going to advising sessions for course/graduation requirements related matters rather than to discuss grades. As such, the perspective of our student participants might not represent the students who are at risk, who are the primary target of SE. However, recruiting students who are academically at-risk was challenging: due to the sensitivity of the topic at hand, we could not identify these students, and furthermore, we were unsure of the potential impact these interviews might have on these at-risk students. However, our student participants' perspectives are still valuable because their data is still used and available in SE.

Another potential limitation is that SE was designed for advisors who work with students on probation. However, as more and more advisors gain access to SE, some advisors do not necessarily work with students who are on probation. In this case, SE's initial design philosophy is not consistently aligned with how advisors perceive and use it. Thus, some of our advisor participants' perspectives on SE might not represent those of advisors who work with at-risk students. Yet, those advisors still access and use SE in interactions with other students.

## 6.2 Insights for Learning Analytics Deployment

Building on our findings from stakeholders' perspectives on SE, we identify insights for the design and deployment of learning analytics systems.

There are many stakeholders becoming connected to the data and with each other due to the existence of learning analytics systems. Thus, regardless of the role they play, it is important to involve these different stakeholders in the design and deployment process of learning analytics systems, as also noted by Prinsloo and Slate. This should go beyond informing them how the learners' data will be used, and include actively learning about their perspectives of the potential impact the system may have on the different stakeholders and the learning environment.

Learning analytics systems might be designed for different purposes and users, but the core stakeholders always include the students because they provide learners' data and they are the intended target that learning analytics strives to improve learning experiences for. In addition, instructors should also be considered because their inputs and involvement affect whether the learners' data can be accurately presented in systems such as SE. Thus, it is critical to be transparent about how learners' data is used with both students and instructors so that they can provide feedback and assist with improvements.

Data accuracy is essential for a learning analytics solution to effectively provide users with sufficient information and fully maximize its features and functions. In order for stakeholders to establish trust and monitor the accuracy of the data in case of uncertainties, the data source, credibility, and reliability need to be properly communicated during the design and deployment process of a learning analytics system.

While learning analytics systems are designed with good intentions, more consideration need to be placed on the privacy and ethical use of learners' data. Being transparent with students about the use, source, and management of the data is important to foster trust relationships between the students and the institution. Surreptitious use of student data without accommodating their needs and rights of the data can be inconsiderate and unethical, as well as negatively affect students' perceptions of their educational institution.

# 7. CONCLUSIONS

Our multi-stakeholder analysis of an early warning system reveals different stakeholders' perceptions, attitudes, and expectations toward the ethics and impact of the access, use, and analysis of learners' data, identifying both tensions and agreements among stakeholders. Our findings can inform the design of learning analytics systems to better involve, integrate, and engage all stakeholders. We advocate for carefully considering alternative use cases and edge-case scenarios in the research and design stage of such applications. Software development processes for learning analytics solutions would likely benefit from incorporating perspectives of the diverse set of users (both direct and indirect) of the systems being developed. Depending on the context, additional stakeholders' perspectives, such as instructors and administrators, should also be considered.

# 8. ACKNOWLEDGEMENTS

This research was partially funded by the University of Michigan's Office of Academic Innovation and School of Information.

# 9. REFERENCES

[1] K. E. Arnold and M. D. Pistilli. Course signals at purdue: Using learning analytics to increase student success. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge*, LAK '12, pages 267–270, New York, NY, USA, 2012. ACM.

[2] R. Baker and G. Siemens. Educational data mining and learning analytics. In R. K. Sawyer, editor, *The Cambridge Handbook of the Learning Sciences*, chapter 13, pages 253–274. Cambridge University Press, 2nd Edition, 2014.

[3] D. Clow. An overview of learning analytics. *Teaching in Higher Education*, 18(6):683–695, 2013.

[4] S. Dawson, D. Gašević, G. Siemens, and S. Joksimovic. Current state and future trends: A citation network analysis of the learning analytics field. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, LAK '14, pages 231–240, New York, NY, USA, 2014. ACM.

[5] M. M. Eliophotou, T. D. Geronimo, and P. Gibbs. *Using data to improve higher education: research, policy and practice*, volume 29. Sense Publishers, Rotterdam.

[6] S. Lonn and S. D. Teasley. Student explorer: A tool for supporting academic advising at scale. In *Proceedings of the First ACM Conference on Learning @ Scale*, L@S '14, pages 175–176. ACM, 2014.

[7] L. P. Macfadyen and S. Dawson. Mining lms data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2):588–599, 2010.

[8] L. P. Macfadyen and S. Dawson. Mining lms data to develop an early warning system for educators: A proof of concept. *Computers & Education*, 54(2):588 – 599, 2010.

[9] A. G. Picciano. The evolution of big data and learning analytics in american higher education. *Journal of Asynchronous Learning Networks*, 16(3):9–20, 2012.

[10] P. Prinsloo and S. Slade. Student data privacy and institutional accountability in an age of surveillance. In *Using Data to Improve Higher Education*, pages 197–214. Springer, 2014.

[11] P. Prinsloo and S. Slade. Student privacy self-management: implications for learning analytics. In *Proceedings of the fifth international conference on learning analytics and knowledge*, pages 83–92. ACM, 2015.

[12] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. D. Kloos. Alas-ka: A learning analytics extension for better understanding the learning process in the khan academy platform. *Computers in Human Behavior*, 47:139 – 148, 2015. Learning Analytics, Educational Data Mining and data-driven Educational Decision Making.