OARS: Exploring Instructor Analytics for Online Learning

Jonathan Bassen, Iris Howley, Ethan Fast, John Mitchell, Candace Thille
Stanford University, Williams College
{jbassen, ethaen, jcm, cthille}@stanford.edu, iris@cs.williams.edu

ABSTRACT
Learning analytics systems have the potential to bring enormous value to online education. Unfortunately, many instructors and platforms do not adequately leverage learning analytics in their courses today. In this paper, we report on the value of these systems from the perspective of course instructors. We study these ideas through OARS, a modular and real-time learning analytics system that we deployed across more than ten online courses with tens of thousands of learners. We leverage this system as a starting point for semi-structured interviews with a diverse set of instructors. Our study suggests new design goals for learning analytics systems, the importance of real-time analytics to many instructors, and the value of flexibility in data selection and aggregation for an instructor when working with an analytics system.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation

Author Keywords
Learning analytics; real-time systems; instructor-centered design.

INTRODUCTION
Systems for learning analytics can dramatically improve the educational experience of students and instructors by measuring, collecting, and reporting on learner data. Prior work in education has shown that learning analytics systems can power many useful course tools, such as interventions targeted at struggling learners [11], analytical methods that help instructors evaluate the coverage and quality of course materials [25], and intelligent tutoring systems that provide automated learning support [28].

Despite these successes, however, instructors often do not take advantage of learning analytics systems in their online courses. Recent suggestions posit multiple explanations for this phenomenon. At the instructor level, these include a need for predictive models customized to the course and context [10]. Learning analytics that power automated interventions are time consuming to author and brittle to changes in course content and learner populations [1]. Further, many analytical measures are difficult to interpret or entirely irrelevant [27]. Motivated by these issues, this paper explores a broad set of challenges and opportunities for learning analytics systems.

To aid this exploration we built and deployed the Open Analytics Research Service (OARS), a modular real-time analytics framework that has now run across more than 10 online courses and 15,000 students. OARS is an extensible system, designed to support a variety of skill models, learning models and visualizations. OARS receives live event streams from independent learning platforms: these events are processed by the learning models to update a variety of analyses, and those analyses are made accessible through visualizations. The learning models and visualizations also draw from the skill models, which are created and adjusted independently. All learner data is stored by anonymous ID, and OARS users must authenticate through their learning platform to access and deanonymize their visualizations.

We leveraged OARS as a starting point for semi-structured interviews with instructors across a diverse set of courses, from statistics to linguistics, to understand their perception of analytics systems in practice. We framed these discussions around three research questions, targeted to inform the design of future systems and maximize their impact on education:

R1: What do instructors hope to gain from using learning analytics systems?

R2: What value do learning analytics currently provide to instructors?

R3: What challenges do learning analytics systems face in bringing value to instructors?

Beyond these interviews, OARS also contributes several technical innovations in learning analytics: a strategy for the real-time transfer of learner data, measures to ensure data security, and an extensible architecture for real-time models and visualizations. While our interview study explores the design space of learning analytics systems more broadly, we also capture useful feedback about OARS specifically that can be applied to future iterations on the system.

Common themes from our interviews surfaced evidence that supports many existing ideas in the learning analytics and information visualization literature, as well as several new findings that are important for the design of future analytics systems. In particular, we found that instructors who use online courses to support in-person interactions with learners need analytics that are truly real-time to support such interactions. We reaffirmed that learning objectives and skill labels are very helpful for instructional processes, but found that

This article was published in the Proceedings of the Fifth Annual ACM Conference on Learning at Scale.

instructors would benefit from content visualization tools to assist them with the process of curating their skill labels. Finally, instructors require flexibility in data selection and aggregation to effectively engage with the outputs from learning models.

BACKGROUND

The interdisciplinary literature on teaching and learning, combined with advances in student modeling, provide a foundation for the OARS framework. As an example, educational psychologists have identified primary factors, such as motivation [16], cognitive-affective state [2], and growth mindset [9] that greatly influence learning. Cognitive scientists have proposed models of human understanding that categorize knowledge into discrete sets of skills that are either known or unknown to a learner [5]. Educational researchers have successfully used skill labels to categorize learning materials, but have demonstrated they are not readily transferable across learner populations; differing exposure to related materials predicates differing levels of skill granularity on various topics [22], [15]. With so many lenses available to focus on a course, it is no wonder no single learning framework has managed to incorporate a significant proportion.

Categorizing and Improving Course Materials

To evaluate the efficacy of course materials, it is helpful to determine the skills and learning objectives of a course; this provides useful scaffolding in determining how well the educational items support the intended knowledge transfer. We define learning objectives as statements describing the measurable expertise learners should possess upon completion of the course, and we define skills as the smallest actionable units of knowledge required to answer questions pertaining to these learning objectives. Together, these form a hierarchy: each educational item is linked with the skills it teaches and/or tests, and each skill is matched to the learning objectives it supports.

The gold standard for assessing skills and learning objectives encapsulated by a course is cognitive task analysis (CTA) [22]. Analytical models, such as Q-matrices, have been developed that approximate CTA by correlating learner performance across assessment items with skill labels for those same items [26]. Unfortunately, these methods assume static learner knowledge state, which is atypical of most online learning environments where the aim is to transfer knowledge over time.

Since CTA is extremely time and labor intensive, a combination of analytical models and human insight can be a more scalable solution. Instructors can draw from their own experiences to create a hypothesized set of skills and learning objectives. These labels can be refined when learners engage with labeled course materials; for example, if two consecutive assessment items with the same skill labels produce significantly different scores, the conceptual difference that's stumping many students on one and not the other is often an unacknowledged or new skill [22]. The OARS system supports this approach.

Modeling Learner Knowledge

When deployed appropriately, real-time learning analytics amplify the explanatory power of assessment materials and provide cues for targeted interventions. Real-time models of learner knowledge leverage interactions with educational items to reflect what learners "know" and predict which assessments they can respond to correctly at a given point in time.

Knowledge tracing algorithms – a class of real-time models that predict learner performance on assessment activities – are often used to enhance online educational environments [17], [28]. To deliver real-time analyses, knowledge tracing models must either be fit to data from previous learner interactions or set to reasonable default estimates [5]. The most basic knowledge tracing algorithm is the n-in-a-row correct model, where a learner is predicted to have mastered a subject area once they've correctly answered several closely related problems back-to-back; from that point forward, the learner is predicted to correctly answer all related problems [1].

Bayesian Knowledge Tracing (BKT) is perhaps the most popular knowledge tracing model among learning scientists, and has been employed by many intelligent tutoring systems and course analysis tools, including [17] and [28]. BKT is driven by a Hidden Markov Model where learner skill mastery is the binary latent variable of interest. Each skill is either "mastered" or "unmastered" for a given student at a given point in time. It is parameterized by the probability a learner knows each skill already, as well as the probabilities of: mastering a skill after engaging with a learning activity, slipping when applying a mastered skill, and guessing correctly on an assessment of an unmastered skill.

Most other real-time models of learner knowledge do not assume binary mastery or non-mastery with respect to individual skills. Similar to BKT, Performance Factors Analysis (PFA) uses explicit skill labels to model knowledge, however the knowledge state for each skill is held as a non-binary score of competency [20]. The standard PFA model has been shown to perform similarly to BKT with cases for slightly better and worse performance [20]. Deep knowledge Tracing (DKT) uses long-short term memory networks (LSTM) to predict learner knowledge state [21] and outperforms BKT when data is abundant. Though skill labels can be used as part of DKT's feature set, learner knowledge with respect to these skills is modeled implicitly and is therefore more difficult to interpret.

Because of its popularity and versatility, we used BKT as the primary analytical model for our first release of OARS. Intelligent tutors have used BKT as a stopping criteria for learners engaging with problems in a specific skill area. BKT mastery estimates have also been used to predict performance on post-tests in online courses that do not employ mastery learning. However, if a knowledge tracing model is used to direct interventions, it may lose this predictive power [1].

RELATED WORK

OARS was designed as a general-case, modular, extensible system for online learning analytics. In this section we describe competing learning analytics system architectures, and instructor-centered design studies for such systems, and explain how they helped inform our design decisions for OARS.

Architecting Learning Analytics Frameworks

Architectural decisions constrain the kinds of analytical models, interactive experiences, and educational contexts where a learning analytics system is useful. The Society for Learning Analytics Researchers (SOLAR) proposed three criteria for useful learning analytics platforms: employing transparent processes and technologies; ensuring modular and extensible development; and allowing for many use cases across platforms and needs [24]. The creators of moocRP [19] have added learner privacy to this list of criteria while [7] adds stipulations for usability and meaningfulness of analytical measures. We were mindful of these issues while creating OARS, and reflect on them in our evaluation.

Several approaches have been taken to create learning analytics systems that are reusable across online learning environments. PSLC DataShop provides a large assortment of analytical tools for anyone looking to do learning analyses or educational data mining on a course dataset [14]. Though powerful, it requires users to export data from their learning platform and import it to the analytics framework, creating extra friction for instructors and precluding real-time analyses. Works such as [29] describe open API standards that can be leveraged for specific kinds of reusable learning analyses, but fail to address multi-platform support, user authentication or learner data security. Typically, the analytics platform polls the learning platform API, and these queries can be computationally expensive for large numbers of learners when paired with complicated learning analyses. The for-profit Learnosity system [18] allows instructors to create educational activities on a single service that can serve these activities to various platforms within an iframe and stream back the results; it also enables learning researchers to build and deploy analytical models that feed off the collected data. Unfortunately, this framework cannot capture learner data from the educational activities that it does not supply and it does not provide persistent storage for analytical results, so it is unable to support advanced analytics whose computations would be too time consuming to recompute with every page load. The moocRP [19] system for learning analytics connects to data exports maintained and managed by individual academic institutions; though moocRP is platform-independent, secure and extensible, this data is not available in real-time.

OARS leverages a lightweight, extensible OAuth API to receive learner data from separate online learning platforms, allowing for real-time analyses while preserving platform privacy controls. In this way, OARS takes a computationally-processed approach to learner data rather than recomputing analyses on each page load, and it is better equipped to handle complex learning analyses.

Designing and Evaluating Learning Analyses

By articulating learning objectives, relating educational content to skills and reviewing BKT analyses, instructors can evaluate their online course materials and track learner mastery over time. However, the packaging and presentation of these analytical practices is likely to affect their usefulness, and there are many learning analyses that could help instructors glean more from their courses. Previous work provides some grounding for the choice and presentation of learning analyses, but, regrettably, most studies of instructor analytics rely on satisfaction surveys and feature usage as their main mode of evaluation, as in [3].

One of the common ways in which learning analytics systems (OARS included) provide for the reporting of data about learners and their contexts, is through information visualization. Infovis is used to support intuitive data exploration without requiring an understanding of complex mathematical models [12]. Our goal for OARS data visualization aligns with the three steps of the visual exploration paradigm in [12]: 1) provide an overview of the data, 2) identifying interesting subsets of data, and 3) support drilling-down into the details.

Prior work has surveyed course instructors to learn about their interests and needs with respect to learning analytics [6], [8]. In these studies, instructors expressed interest in: qualitative evaluation of course content and experiences as expressed by learners, quantitative measures of content use, differentiation between groups of learners, distinctions between educational items, learner performance and outcomes, as well as the relationships between these different measures.

However, not all analytical measures are equally meaningful, and potential measures should be carefully scrutinized before being included. [23] finds that regimented adherence to mastery learning (where learners must repeatedly demonstrate competence with each skill before moving on from a module or course) significantly improved outcomes in a blended online/in-person learning environment. [4] recommends that learning analytics designers prioritize learner progress with respect to instructor's learning goals, and be wary of inappropriate and misleading metrics. Further, [4] argues learning analytics for instructors should support iterative hypothesis testing to improve course materials. The challenge is to supply analytics that are generalizable across courses and learning environments, without sacrificing predictive power. To this end, the first release of OARS featured skill tagging and knowledge tracing capabilities, which are broadly applicable and can be used to evaluate course materials and track learner mastery.

OARS DEPLOYMENT

Through the breadth and scale of OARS's deployment, we aimed to to better understand the motivations and needs of online course instructors. In this section, we provide additional detail about the set of classes and learners supported by OARS, our information gathering process, and the system's various stages of deployment.

Courses and Students

OARS has provided learning analytics to more than ten online courses with thousands of students. Table 1 provides summary data about the courses, which covered a breadth of content: from statistics and linguistics to philanthropic strategy and remedial math. Two of the courses ran on our institution's public Open EdX platform, while the remaining four ran on our institution's student-oriented platform. Learner populations similarly ranged from students at private research universities to enrollees of open online courses.

Instructor	*Context*	*Learner Population*	*Learners*	*Sections*
A1	online	open and free	~12,530	1
A2	flipped	R1, private	~45	1
U1	online	R3, public	~1,590	1
U2	online	R1, private	~1,950	1
U3	online	R1, private	~325	1
U4	flipped	liberal arts	~75	3
U5	flipped	R1, public	~215	1
U6	flipped	R3, public	~150	2

Table 1. Course information for the instructors who used OARS. A1 and A2 were author instructors. U1-U6 were instructors that we recruited to use OARS with their courses. R1-R3 denotes the Carnegie classification of the institution, when appropriate. Learner enrollments were approximated by the number of registered users at the conclusion of the course.

Information Gathering

Development of the Open Analytics Research Service was initially motivated by our own instructional needs. We hoped to reach a wider audience, benefit from the large community of open source developers, and explore new formats for educational activities. Our learning objectives, skills and course content were developed using CTA and were further refined using post-hoc analyses following learner engagement with the materials. Critically, we needed a real-time analytics solution that was platform-independent.

Before developing OARS, we first spoke with instructors and instructional designers for our local statistics course to distill their data needs. We intended to deploy this course as both a stand-alone, self-directed online learning experience, and also as a mixed offline-online course where learners engage with new material and practice problems in-person and then complete many more online problems outside of class.

Given these scenarios, we discovered two primary needs. First, instructional designers wanted to visualize the relationships between the learning objectives, skills and problems, so that they could see learner progress with respect to these learning objectives and skills. The goal would be to identify struggling learners and skills where a large number of learners were not showing mastery. Second, instructors needed access to learner performance on the individual problems – to identify harder problems and go over them in greater detail. We, the framework developers, also wanted to create an analytics framework that was protective of learner privacy, able to accommodate a variety of analytical models and visualizations, and open-sourced for future development, inspection and deployment.

Release Stages

For our initial OARS deployment, we implemented a minimal set of models and visualizations to evaluate course content and to assess learner progress. We built a spreadsheet-based service for mapping learning objectives to skills and skills to course problems, and we constructed two BKT algorithms to predict learner mastery across the articulated skills: one which made predictions based on learner's first attempts, the other which tracked every problem attempt.

After this short trial, we recruited six instructors from five institutions of higher education who were interested in improving their courses through the analytical process supported by OARS. These courses are described in Table 1.

Developing Skill and Concept Labels

CTA is ideal for categorizing and improving course materials, and bringing attention to the course skill set. However, this process is time consuming, requires several volunteers, and still needs adjusting when the content is delivered to a new learner population. The instructors who were interested in OARS integration were not prepared to go through a full round of CTA before running their courses; instead, we had conversations about learning objectives and skills, and answered questions for those instructors creating new content and mapping skills onto course problems.

OARS ARCHITECTURE

In designing OARS, we confronted many challenges faced by modern analytics frameworks. The OARS architecture addresses issues of data security, extensibility, portability and real-time accessibility in a manner that advances the state of the art and better serves online instructors' needs.

Data Flow And Storage (Figure 1)

To ensure extensibility, OARS employs MongoDB, a popular NoSQL database, for its internal storage; this allows the data in each table to take on many types (structural forms) instead of requiring a new table for each additional data type that gets added to the system. This database allows for five primary modes of storage: registered courses and learner registrations, structural course data, raw learner data, models on course content, and learner models. The registration listings enumerate the learning platforms, courses associated with each platform, and the learners enrolled in each course. The structural data lists out the course modules, problems, any text associated with the problems and mappings between modules and problems. The raw learner data is intended for all learner interactions with course material, including attempted problems, submitted answers, and scores. The models of course content are extensible, but currently contain learning objectives, skills, mappings between the learning objectives and skills, and mappings between the skills and problems. Finally, the learner models encapsulate any information stored about individual learners that goes beyond their raw interaction data, such as predicted knowledge state with respect to a given analytical model; learner models that pertain to the entire course population can be stored using an "all learners" qualifier instead of a specific learner ID, though we have not yet encountered course-wide models which were not readily derived from the individual learner models.

The course IDs and learner enrollments were the easiest data to obtain. Our institution hosts two separate versions of EdX – one instance intended for student learning and one instance for externally directed courseware. The OARS administrator adds the course ID and platform name into the system. EdX supports OAuth-protected enrollment APIs to access the list of instructors and learners associated with a given course ID. OARS uses each instance of EdX as an OAuth provider, to

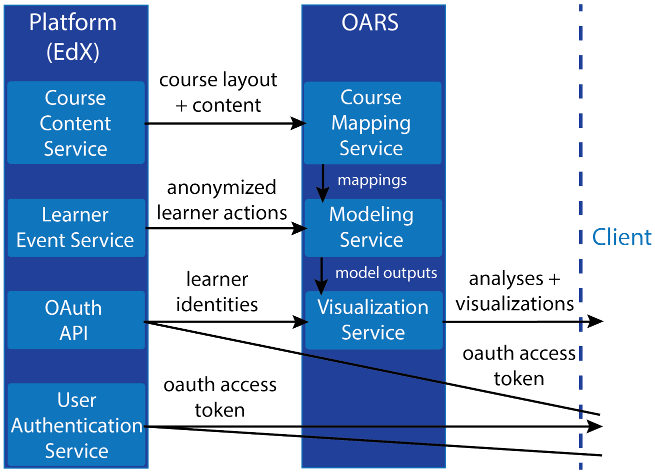

Figure 1. The OARS system is in constant communication with every linked-up learning platform, updating its analyses in real time to support immediate user inquiry. This diagram shows the data flows between OARS, the connected learning platform, and the user's browser; it also depicts the flow of data between the OARS services.

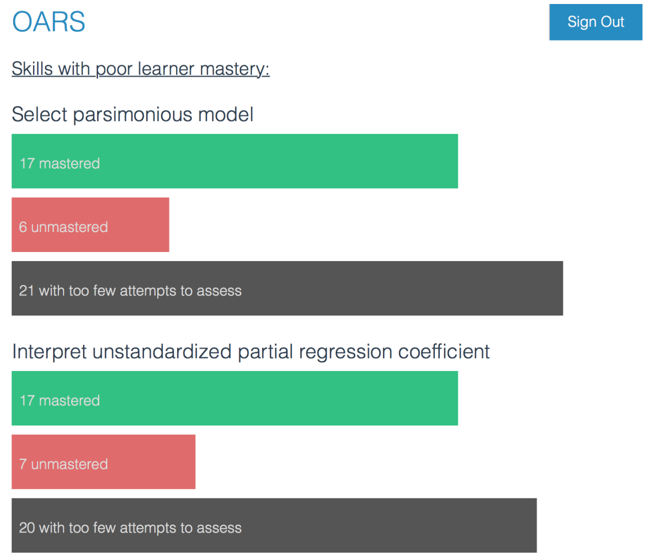

Figure 2. The course landing page directs instructor attention to recently encountered skills where learners are performing poorly.

authenticate and log in its users; OARS then takes the resulting OAuth credentials to query the enrollment API, retrieving and updating the course roster on their behalf from the platform they have authenticated with.

Obtaining the structural course data and course content models requires direct user input. EdX has an API for accessing course content, but it does not list it in order of module or provide the text associated with each piece of content. To get around this, an instructor and/or course developer provides the OARS administrator with a spreadsheet containing ordered entries for each item in their course, along with the course module, skills and learning objectives associated with it. (The item IDs are available to the instructor and/or course developer when they author content on EdX.) The OARS administrator then uploads this spreadsheet to extract the course structure and populate the course content models for skills and learning objectives. To retrieve the text corresponding to each educational item, we created a nightly task within EdX that would collect all the text for all registered courses and send it to OARS.

After exploring several ways of obtaining raw learner interaction data from the EdX platform, we determined it was best to transmit a real-time stream of learner interactions directly to OARS via secure HTTPS-encrypted connections. Learner interactions within EdX are written to their logs, but not exhaustively stored to their internal databases – precluding API access. Nightly log transfers, though feasible, would increase the latency of learner modeling from a number of seconds to a full day, which we deemed unacceptable. In collaboration with our institution's EdX platform team, we extended EdX to scrape real-time learner events from the platform logs, and transmit these events to OARS via a REST API call – for all events corresponding to one of the registered OARS courses.

The OARS REST event API, authentication service, EdX query engine, and web interface are all built on Tornado, an asynchronous web framework for Python. Asynchronous web frameworks are better equipped to handle many concurrent events, such as HTTP requests, database reads and writes. Tornado is able to leverage multiple server cores to allow for even greater scalability.

Real-Time Models and Visualizations

OARS was designed to efficiently host a number of learning models. Each model hosted on the analytics framework receives a dedicated private virtual machine (VM), read access to all non-administrator database documents and limited write access to store its analyses. To prevent excessively large database queries, we impose minor limitations on the size and complexity of allowable queries; this requires models to cache intermediate results and course maps on their private VMs, consuming fewer system-wide resources. These analytical models poll for new and relevant learner events, iteratively updating the model state for each learner. The models also poll for updates to course content to trigger internal structural refreshes as needed.

Our analytics framework enables web-based visualizations in the form of independent JavaScript documents and supporting MongoDB queries. Each JavaScript document generates a single-page visualization, and gets assigned a context-dependent variable path for user access. Mongo queries aggregate and restructure the results from the OARS analytical models, and share this data by appending it to the top of the JavaScript document when it is sent to the user's browser for rendering. In this way, each visualization can flexibly link to any other, new visualizations can be added at any time, and model data gets updated on each page load.

At the time of its release, the OARS system provided three hierarchically linked visual representations of the course content and learner progress. Each instructor was met with a list of the course learning objectives, from which they could drill down to see the skills related to an individual objective, and

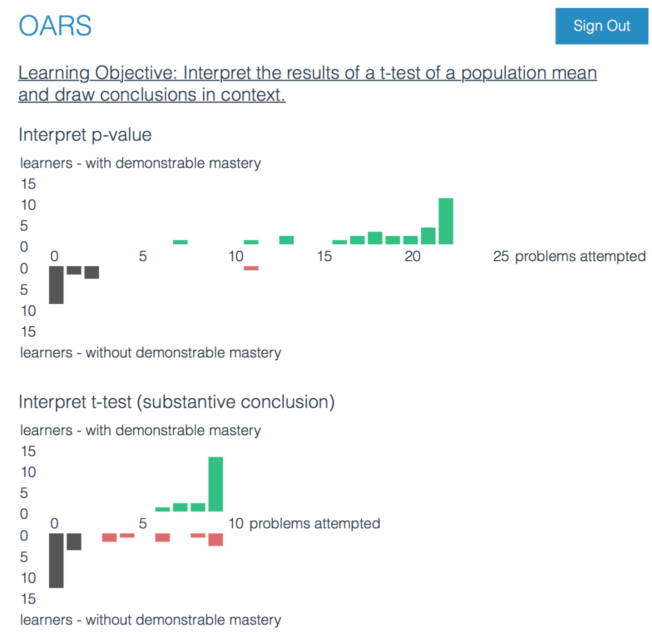

Figure 3. Each skill visualization plots the number of learners who have mastered a skill against the number of problems they have attempted. In the second case, several learners who have attempted many problems have failed to achieve mastery over the skill: "Interpret t-test."

finally click on a skill to see the problems associated with it. The idea behind this design was to direct instructor attention towards learner progress in terms of skill mastery and stated learning objectives rather than individual problem scores. The problems pages showed every problem associated with a skill, the number of learners who attempted each problem, and the number who got it right on their first and final attempt. The skills pages visualized learner mastery and non-mastery, in relation to the number of problems each of the course's learners had attempted involving said skill (Figure 3). The learning objectives pages visualized the number of learners who had demonstrated mastery on all the associated skills (Figure 4). To bring attention to skills with low levels of learner mastery, the landing page was expanded to display the three skills where learners had encountered enough problems to demonstrate mastery according to the BKT algorithm but were failing to show proficiency (Figure 2).

Authentication and Security

Though learning analytics necessitate access to learner data, our system makes every effort to safeguard learner privacy. With a platform-dependent OAuth login, we prohibit access to course data for any party other than registered instructors within the online learning environment. By verifying their enrollment status through the EdX enrollment API, we prevent instructors from accessing the learner data for courses where they have no verifiable affiliation. Our framework's servers communicate via encrypted web protocols over a private network, the database servers are disabled from receiving externally-initiated internet traffic, and we employ firewalls and HTTP rate limiting to prevent common forms of server intrusion. As an added precaution, we never store personally identifiable information (PII) on OARS and instead use

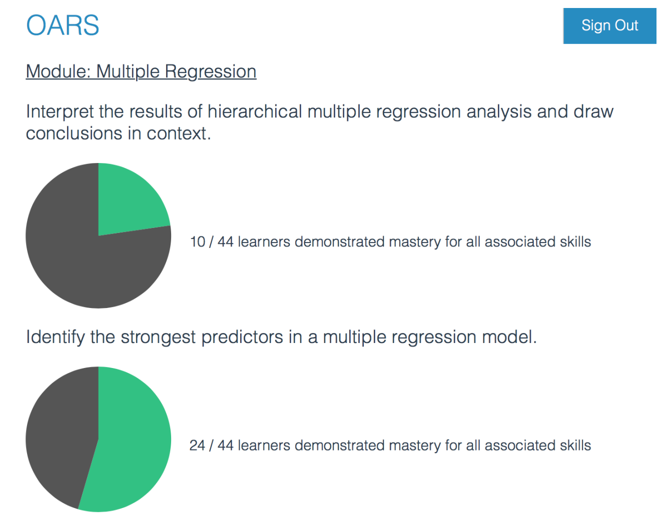

Figure 4. These visualizations show the proportion of learners who have mastered all of the skills associated with each learning objective.

externally-generated EdX learner IDs. Learner names are matched with learner IDs in process memory, as part of the instructor-authenticated enrollment API query. In contrast, existing analytics frameworks either omit PII – preventing analytics users from identifying individual learners – or copy PII onto internal databases where it must be carefully safeguarded.

EVALUATION

Our work with OARS provided a starting point for interviews with instructors across a diverse set of courses. These interviews were aimed at uncovering the challenges and opportunities for learning analytics in online education, with the goal of informing the design of future analytics systems.

In this section, we report on the structure and execution of these interviews, centered around our three research questions. We further discuss how instructors reacted to specific design decisions made by OARS, and how these reactions can provide value for future iterations of the system.

Method

We conducted semi-structured interviews with twelve instructors who ran fully online or partially online courses: the six instructors who had used OARS in conjunction with their courses, and six additional instructors who had previous experience with learning analytics but had not used OARS. We began by gathering information on the instructors, their courses, course materials and educational environments. All of the instructors had at least five years of teaching experience, and most had more than fifteen. The instructors who used OARS taught at public and private research universities, as well as institutions that focused on undergraduate education. The set of instructors who had not used OARS were similarly representative, with added diversity from two community college instructors. This group of nonusers contributed additional diversity in terms of the courses they taught; though there were no additional linguistics or pre-calculus instructors, we spoke with instructors for online courses on philosophy, physics, and statistics for the social sciences.

We discovered there was further diversity in the ways that instructors developed course content. Two of the instructors using OARS assembled their own learning materials from scratch, in conjunction with self-designed learning objectives and skills; one instructor borrowed learning materials but made significant adjustments before using them; the remaining three made less significant adjustments to the skill-labeled learning materials they used. The six instructors who did not use OARS followed similar practices: three designed their own courses using learning objectives, alongside skills or "concepts" (it was unclear how closely the concepts they described resembled our notion of skills); the remaining three non-users made use of available online learning materials.

The remainder of our interviews were focused on the utility of OARS, and the gaps between instructors' analytical needs and the analyses provided by existing analytics frameworks such as OARS. We asked the instructors to list out all the tasks they wished learning analytics could assist them with; the intention was to get a first pass at R1 in the absence of any known artifact. After talking through these wish lists, we went over OARS visualizations with non-users to get their feedback, and spoke with OARS-empowered instructors about their experiences using the platform; this discussion elicited evaluations of OARS and previous learning analytics frameworks, providing insights relevant to R2 and R1. Finally, we would return to discussing instructor's analytical wish lists, to see what else might be missing from existing learning analytics frameworks.

The interviews were recorded, reviewed and discussed by two researchers who were familiar with learning design and trained in qualitative HCI methods. These researchers took notes on each interview, compared interpretations, and grouped the responses into broad themes. We report on these responses here through the lens of our three research questions.

R1: What do instructors hope to gain from analytics?

We found that all of the instructors wanted to identify concepts that learners were struggling with. Instructors with in-person learner interactions wanted to use this information to review or emphasize more difficult material; the instructors with smaller courses wanted to know which learners were struggling with each concept to attempt individual interventions. Instructors running online courses wanted this feedback so that they could provide more content relating to the most difficult concepts in their courses. Those who were publishing their course materials online for the first time wanted to revise their materials in the areas with the lowest mastery rates. Conversely, instructors who had taught their course for several years appeared to trust their course materials more; though they would spend more time on these concepts, they were more likely to assert that these concepts were just harder to master.

The three instructors who were in the process of deploying new courses expressed interest in analytics to aid with course design. All three described a tabular process of listing out the concepts they wanted to go over and identifying educational activities they wanted to pair with these concepts. Apparently, it was difficult to evaluate the amount of coverage given to each concept when looking at a spreadsheet because the educational items were not necessarily ordered by concept, the items regularly covered more than one concept, and the items were of variable length.

Half of the instructors interviewed expressed regret that they did not know more about their learner populations. Three explicitly called out times that learners were going through significant life events outside of class, such as the death of a family member or birth of a child, which significantly affected course performance and wondered how many times learners had neglected to share this information. Other instructors were interested in obtaining sociodemographic information, prior coursework, and reasons for learner enrollment. Six instructors brought up psychological factors including motivation, growth mindset, and anxiety, which they believe has significantly impacted learner performance in their courses.

Seven of the instructors mentioned they wanted a better way to grade and analyze answers for free-form questions; they found these types of problems to be valuable both for learning and assessing. Four stated they did not think analytics technology was adequately advanced to meet these needs. Three mentioned possible solutions involving rubrics, peer grading, and/or natural language processing.

The instructors who had in-person interactions with students all indicated that real-time access to learner data was a firm requirement. Seven mentioned that they would like to review the results of the previous day's assignment. One instructor said: "The way that [the last online learning system operated], they were running these reports 24, 48, 72 hours late, which was useless. ... [The platform developers] were saying most [instructors] don't use this technology the way that you're using it. ... They should be. If they're not using it that way, they ought to to be using it that way." Four instructors mentioned that they wanted to keep tabs on online learner activity over the course of a week to track progress; some wanted to have this information to help discourage procrastination, while others used this as a secondary indicator that learners were getting stuck. Real-time data appeared less strict for the three instructors whose online courses did not have strict deadlines.

R2: What value do analytics provide to instructors?

Instructors strongly valued the idea of skill labels, finding that these made it easier to organize and think about the content. The two instructors who used OARS while launching new online courses said that the system helped them to visualize the course content and identify skills that had fewer related problems than they would have liked. The instructors who augmented their course materials said that the number of problems available per skill greatly influenced which of the problems they cut and added. The two instructors who designed their own courses and used the skill labels when naming their educational items found that learners had a better vocabulary to describe what they were struggling with.

Most instructors also found value in measures of learner mastery. The instructors that used OARS regularly reported using learner mastery predictions to decide what to review in class, and when it would be beneficial to go over worked examples of specific problems. A common process for using OARS was described as follows: "So the dashboard gives me an idea of whether they are keeping up with the reading and how they are doing with their first attempt on the questions and it gives me a way to plan what I'm going to discuss in class. I might copy certain excerpts from what they seem to be struggling with, then we can go over those passages in class." Instructors use the OARS system for more than simply checking participation, but also as a means of tracking the class' progress and tailoring the pedagogical approach to student challenges. One instructor said, "It broke the ice when I told students two-thirds of them were struggling with a skill. They were no longer embarrassed to ask questions." However, one instructor whose learners showed unusually high mastery across the skill set did not find these mastery predictions to be particularly useful, stating that they would rather focus on the individual problems learners failed to solve than addressing the skill more generally.

To contextualize instructional needs, it was interesting to hear the diverse factors that led each of the instructors to their current online learning environment. Eight of them were primarily attracted by the range of learning activities supported on their current platform. Two who were using our statistics course materials and one instructor who was re-purposing a colleague's content said they simply chose to use the platform where this content was readily available to them. Two instructors said that it was most important that they be able to extract raw learner data and run their own quantitative analyses. Four instructors cited institutional support for their platform of choice, as well as access to platform-specific tech support.

R3: What challenges do learning analytics face?

Most instructors who had in-person interactions with learners said that while helpful, they did not think mastery predictions based on out-of-class problems was necessarily sufficient for determining learner knowledge. One instructor commented: "I don't have a complete reliance on all the information that's in OARS because I don't know if students are working together – if it's showing me what students can do together or if its individual. And since the environment that I create is so collaborative, I want them to work together. ... It's a bit difficult for me to place heavy reliance on what I see in that and I tend to place more reliance on what I see in the class and ... low stress environments." The three instructors who had no in-person interactions with learners said they found these measures of mastery preferable to single high-stakes events.

Similarly, eight of the instructors stated that they would have liked to explore deeper connections between various groupings of learners and their levels of mastery. This is difficult to achieve in most learning analytics systems, including OARS. Examples of groupings included distinct course populations within on-campus course sections, learners who had failed to master a particular skill, learners who had given a certain response to a problem, and learners who had interacted with the system over a particular window of time. (The latter was a serious sore spot for one fully online self-directed course where learners could enroll at any time.)

Skill map creation is another big issue for analytics systems. All of the instructors who submitted skill maps for OARS preferred a user interface for creating these maps rather than submitting a spreadsheet. They would also like the course problems and associated text to be available within OARS while performing this task. Several instructors identified errors in their spreadsheets shortly after they had been uploaded.

Another big point of frustration for some instructors came from the difficulty that learning analytics systems have in capturing a variety of learner answers. All of the instructors who used OARS requested either that we add visualizations to show the number of users who had selected a particular answer choice to a problem, or that we add greater support for open-ended responses which get graded at a later time – often according to a rubric.

DISCUSSION

OARS's mapping, modeling and visualization functionalities enhanced users' understanding of course content and learner knowledge over time. However, the mastery model was only useful when there was significant variance in learner performance across the skills, and when the problem format was amenable to automated grading. These results suggest that knowledge tracing analyses should be made available to all instructors whose course problems are auto-graded and skill-labeled. If instructors' learning objectives require learner mastery across the skills, it is not necessarily problematic when the majority of learners demonstrate this mastery, even if the resulting models are less actionable for instructors.

Although most instructors found our BKT-driven visualizations useful for organizing time spent in class and adjusting online content, they pushed for several changes to these visualizations. The instructors who engaged with learners in-person wanted a visualization that revealed the predicted skill mastery states for individual learners, all in one place. They thought this information would allow them to more easily identify struggling learners and carry out personalized interventions. Most of the instructors requested additional flexibility in data selection and aggregation, to find correlations and test hypotheses using the outputs from the BKT model. In summary, instructor analytics on learner knowledge state should allow them to inspect predictions across all learners, at the level of individual learners, as constrained by learner features, and as constrained by time frames. There was also consensus among the instructors that analytical results ought to follow the same order as course content, but be primarily organized by skill.

In online courses with in-person interactions, instructors unanimously demanded real-time analytics. These instructors used up-to-date analyses to intervene when learners were not doing their work or appeared to be struggling with content. This suggests that analytics frameworks targeting instructors with direct learner contact should prioritize the speed and ease with which data is transferred and processed, when modeling learner knowledge state.

All of those interviewed thought it was beneficial to identify learning objectives and skills – for developing course materials, organizing courses, and monitoring learner progress. However, there was considerable friction in developing these descriptors, and visual mapping tools were sought to facilitate this process. We do not anticipate any significant difficulty in providing such a tool as a part of a modular analytics service, so long as the associated learning platform enables external access to course materials. Therefore, we would recommend that analytics frameworks include a graphical interface and visualizations to assist with course map development.

The instructors were unified in asking for improved capture of learner response data. Some wanted to see which wrong answers learners were providing, to identify learners' incorrectly understood skills or misconceptions [13]. Others sought more meaningful analyses for free-form learner responses, such as conceptual explanations and long-form calculations. Recently published models that evaluate essay responses [31] and deconstruct learner code [30] present potential solutions for these shortcomings. Our interpretation from instructor feedback is that all learner responses should be made available to instructors in one place, and analyzed according to a set of metrics if possible; also grouped by response when possible.

Though instructors indicated that they would like additional data about their learners, it was not always clear how this data would be beneficial. Previous research has shown that sociodemographic information can be used from the onset of a course to predict which learners will struggle within a course. The same may be true of previous courses taken and significant life events occurring simultaneously. However, it is unclear where to draw the line between protections for learner privacy and support for instructor decision making. Even if certain data are useful, we suggest that this class of information should be revealed at the discretion of the individual learners.

We assert that OARS improves access to learning analyses. Although several online learning platforms offer advanced analytical capabilities, there is typically a single model for learner progress baked into the platform. Instructors are unable to use other models without exporting their course data and loading it into an external analytics framework such as DataShop [14], and this process impedes real-time analysis. Though attempts have been made to improve the portability of online course materials [18], there is a significant time cost to transferring content across platforms. Instructors that use multiple learning platforms for a single course have an especially hard time evaluating learner progress and content without doing some data wrangling; whereas we are currently updating our analytics framework so that it can store and analyze data from multiple platforms under a shared course ID.

Although OARS was designed with instructor analytics in mind, learners could certainly use a modified version of our system in the future. Direct learner engagement with the analytics framework would also make it easier to collect additional metacognitive and affective data pertaining to learner mindset, affect, and motivation. To reduce the friction of using a secondary system, we would likely embed page links to the learning platform to facilitate learner navigation.

CONCLUSION
In this paper, we report on findings from a large scale deployment of OARS, a modular real-time analytics platform for online education. Our work seeks to better understand the design space for learning analytics systems, suggesting and reaffirming many instructor goals such as the importance of real-time analytics and flexibility of analytic tools for data manipulation. By better understanding instructor needs, we hope to empower future learning analytics system to maximize their positive impact both on students and instructions.

ACKNOWLEDGMENTS
Many thanks to David Adams and the engineering staff at the office of the vice provost for teaching and learning for exceedingly generous support with Open EdX. We are grateful for the enthusiasm and patience of the instructors who used OARS and made this research possible. We also thank the volunteers, publications support and staff, who wrote and provided helpful comments on previous versions of this document. This research was supported by NSF award #1443068.

REFERENCES
1. Ryan S Baker. 2016. Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education* 26, 2 (2016), 600–614.
2. Ryan S.J.d. Baker, Sidney K. D'Mello, Ma.Mercedes T. Rodrigo, and Arthur C. Graesser. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners‘ cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* 68, 4 (2010), 223 – 241.
3. Sven Charleer, Jose Luis Santos, Joris Klerkx, and Erik Duval. 2014. Improving teacher awareness through activity, badge and content visualizations. In *International Conference on Web-Based Learning*. Springer, 143–152.
4. Doug Clow. 2012. The Learning Analytics Cycle: Closing the Loop Effectively. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge (LAK '12)*. 134–138.
5. Albert T. Corbett and John R. Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* 4, 4 (01 Dec 1994), 253–278.
6. Anna L Dyckhoff. 2011. Implications for learning analytics tools: A meta-analysis of applied research questions. *International Journal of Computer Information Systems and Industrial Management Applications* 3, 1 (2011), 594–601.
7. Anna Lea Dyckhoff, Dennis Zielke, Mareike Bültmann, Mohamed Amine Chatti, and Ulrik Schroeder. 2012a. Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society* 15, 3 (2012), 58.
8. Anna Lea Dyckhoff, Dennis Zielke, Mareike Bültmann, Mohamed Amine Chatti, and Ulrik Schroeder. 2012b. Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society* 15, 3 (2012), 58.
9. Andrew J Elliot and Carol S Dweck. 2013. *Handbook of competence and motivation*. Guilford Publications.

10. Dragan Gasevic, Shane Dawson, Tim Rogers, and Danijela Gasevic. 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education* 28 (2016), 68 – 84.

11. G. Lopez C. Coleman J. Whitehill, J. Williams and J. Reich. 2015. Beyond prediction: First steps toward automatic intervention in MOOC student stopout. (2015).

12. Daniel A Keim. 2002. Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics* 8, 1 (2002), 1–8.

13. Kenneth R Koedinger, John R Anderson, William H Hadley, and Mary A Mark. 1997. Intelligent tutoring goes to school in the big city. (1997).

14. Kenneth R Koedinger, Ryan SJd Baker, Kyle Cunningham, Alida Skogsholm, Brett Leber, and John Stamper. 2010a. A data repository for the EDM community: The PSLC DataShop. *Handbook of educational data mining* 43 (2010).

15. Kenneth R Koedinger, Albert T Corbett, and Charles Perfetti. 2010b. The knowledge-learning-instruction (KLI) framework: Toward bridging the science-practice chasm to enhance robust student learning. *Cognitive Science* (2010).

16. Steven Lonn, Stephen J Aguilar, and Stephanie D Teasley. 2015. Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior* 47 (2015), 90–97.

17. Marsha Lovett, Oded Meyer, and Candace Thille. 2008. JIME-The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education* 2008, 1 (2008).

18. Learnosity Ltd. 2006-2018. Learnosity. Website. (2006-2018). Retrieved January 21, 2018 from www.learnosity.com.

19. Zachary A Pardos and Kevin Kao. 2015. moocRP: An open-source analytics platform. In *Proceedings of the Second (2015) ACM conference on learning@ scale*. ACM, 103–110.

20. Philip I Pavlik Jr, Hao Cen, and Kenneth R Koedinger. 2009. Performance Factors Analysis–A New Alternative to Knowledge Tracing. *Online Submission* (2009).

21. Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. 2015. Deep Knowledge Tracing. In *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Eds.). Curran Associates, Inc., 505–513.

22. J. vanMarrienboer K. Yates R. Clark, D. Feldon and S. Early. 2008. *Handbook of research on educational communications and technology* (3rd ed.). Chapter Cognitive task analysis for training, 577–593.

23. Steve Ritter, Michael Yudelson, Stephen E Fancsali, and Susan R Berman. 2016. How mastery learning works at scale. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 71–79.

24. George Siemens, Dragan Gasevic, Caroline Haythornthwaite, Shane Peter Dawson, Simonbuckingham Shum, Rebecca Ferguson, Erik Duval, Katrien Verbert, Robert Baker, and others. 2011. Open Learning Analytics: an integrated & modularized platform. (2011).

25. S. Levine E. Brunskill T. Mandel, Y. Liu and Z. Popovic. 2014. Offline policy evaluation across representations with applications to educational games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 1077–1084.

26. Kikumi K Tatsuoka. 1995. Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively diagnostic assessment* (1995), 327–359.

27. Dirk T Tempelaar, Bart Rienties, and Bas Giesbers. 2015. In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior* 47 (2015), 157–167.

28. J. Sewall V. Aleven, B. McLaren and K. Koedinger. 2006. The cognitive tutor authoring tools (CTAT): preliminary evaluation of efficiency gains. In *Intelligent Tutoring Systems (ITS '06)*. 61–70.

29. Andrii Vozniuk, Sten Govaerts, and Denis Gillet. 2013. Towards portable learning analytics dashboards. In *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*. IEEE, 412–416.

30. Lisa Wang, Angela Sy, Larry Liu, and Chris Piech. 2017. Learning to represent student knowledge on programming exercises using deep learning. In *Proceedings of the 10th International Conference on Educational Data Mining; Wuhan, China*. 324–329.

31. Bronwyn Woods, David Adamson, Shayne Miel, and Elijah Mayfield. 2017. Formative Essay Feedback Using Predictive Scoring Models. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 2071–2080.