**Project Completed by:**

EDUCAUSE

Industry and Campus Webinar – In-Depth Design: Data-Driven Design Strategies to Support Learner Outcomes

October 16, 2018

1:00PM – 2:00PM Eastern

**EDUCAUSE**

**Industry and Campus Webinar - In-Depth Design: Data-Driven Design Strategies to Support Learner Outcomes**

**October 16, 2018**

>> Welcome to today's Industry and Campus webinar: In-Depth Design: Data-Driven Design Strategies to Support Learner Outcomes. This is Adam La Faci, Online Event Producer with EDUCAUSE and I'll be your moderator for today. EDUCAUSE is pleased to welcome today's speakers: Kenneth Koedinger and Padraig Nash. I will introduce them in just a moment, but first let me give a brief orientation on our session's learning environment. Our virtual room or learning space is subdivided into several windows. Our presenter's slides are now showing in the presentation window, which is the largest on the screen. The tall window on the left is the chat window, serving as the public chat space for all of us. You can use the chat for comments, to share resources, or to pose questions to our presenters. We will hold Q&A until the end of the presentation; but we encourage you to type your questions into the chat throughout the webinar. If you have any audio issues, click on the link in the lower left-hand corner. And, at any time you can direct a private message to "Technical Help" by clicking in the top-right corner of the Chat Pod. A drop-down menu will appear where you can select "Start Chat With" and "Hosts." The session recording and slides will be archived later today on the EDUCAUSE website. And now, let's turn to today's presentation. Designing for digital learning experiences requires thoughtful, deliberate choices that will effectively lead the learner to new understandings, problem solving and, ultimately, completion. A quality design process should include key beneficiaries of the designs-instructors, students, and subject matter experts-as well as educational data mining (EDM) and learning analytics that allow instructors and other educational providers to test and expand pedagogical content knowledge. In this webinar, the speakers will discuss how solutions that use EDM and learning analytics (LA) can expand practitioners' pedagogical content knowledge and their ability to provide effective feedback to learners. You will leave with a deep knowledge of how EDM and LA are providing insights into how students learn (for example, slowly and incrementally) and how resources using EDM and LA can foster improvements in student learning outcomes and dramatically increase effective learning. We are delighted to be joined today by Padraig Nash. Padraig has worked in the field of education for over twenty years, as a teacher, teaching-artist, researcher, consultant, program director, and designer. He has a PhD in the Learning Sciences from the University of Wisconsin-Madison, where he designed and studied mentoring in virtual internships. He is currently a Learning Engineer at Cengage Learning. We are also joined by Kenneth Koedinger, a professor of Human Computer Interaction and Psychology at Carnegie Mellon University. He has authored over 250 peer-reviewed publications and has been a project investigator on over 45 grants. In 2017, he received the Hillman Professorship of Computer Science and in 2018, he was recognized as a fellow of Cognitive Science. And with that, let's begin today's Industry and Campus webinar: In-Depth Design: Data-Driven Design Strategies to Support Learner Outcomes. Ken, over to you.

>> Great. I am going to share my screen now. So I can continue with the slide deck here. Just one moment. Are we visible?

>> That's perfect, thank you, Ken.

>> Great. So let me start with some broad goals for educational data mining and learning analytics. We want to use these kinds of approaches to understand challenges students may face in their content or cognitive skill development. And also in their learning skill development. And third, with respect to motivation, and in all case, of course, it's not simply about the data, it's ultimately about making differences in the instructional design that can improve student learning. There's been a lot of work in all three of these area, and I wish I had three hours to talk about all of them. But today I'm just going to focus on the first of these, the content and cognitive skill areas. I put a couple references to some review papers that we have written down at the bottom here. On the left is more oriented towards cognitive science psychology, and on the right is more oriented to computer science, A.I.-kind of audience. So digging in here, I'm going to first talk about how data enhances our intuition, sometimes contradicts our intuitions. Let's make a brief comment in the details I give about the role of instructors and then hand it off to Padraig to talk about making this work in practice. So to get us rolling on this, we have a second poll here. And so what I would like you to do in this poll is reflect on which of these three kinds of problems is harder for beginning algebra students in a story problem, a word problem, and an equation. As you process those, you might notice they all have the same numbers and the same answer. And there's six and 66 and 81.90 and how many hours did Ted work? The word problem has no story context, but stated in English starting with some number, multiplied by 6 and then add 66, I get 81.90. What number did I start with in and equation, I think you can see there at the bottom. So the question is, I'll show you some data in a moment of high school students about six or seven months into a high school algebra course. This data we collected in the early '90s. And we will look at the percent correct for problems in these three categories. These being representative problems in those three categories so. Which one is the hardest?

>> Ken --

>> So what are the polling numbers looking like Padraig?

>> The story problem is around 55%. The word problem about 35%. Now down to 32. And the equation at about 9%.

>> About 9% great. So most of you are saying the story problem, your word problem. I would love to hear your explanations and thoughts about that. You could pop those in the chat if you like. And a few are saying equation. I actually recognized one of you who has probably seen this before, I so hope Bruce you're saying equation. Because while math educators say story problems and word problems are the hardest and we surveyed math teachers and confirmed indeed their intuitions were, along with your in fact along with mine when I originally designed this data collection, I was trying to understand why the story problems are so hard, but the surprise was that they're not the hardest. That the story problems and word problems are solved at a higher percent correct rate than the equations. And substantially so. And we actually replicated this numerous times over the years in various different formats. And we showed that when you increase the complexity of the problem on the story side and then matched more complex equation, when the unknown is referenced twice, the relationship does flip. And indeed equations are important and powerful. But for these beginning algebra students, they're actually better able to solve the story in word problem than the equation. And that kind of nominal task analysis we might do that suggests that well, to solve the story, tough translate it to an equation so of course the story has to be harder doesn't work here. And two things are going on. Students are doing better than expected on the story and word problems and they're using informal strategy, they go directly to the math without using the equation. Sometimes they do a generate and test strategy quite efficiency. They do better at those than expected. But the other thing which in the end which is probably more important, the difficulties with equations are more than we might expect. And what we came to understand, they're having trouble reading the equation. You don't think about reading equation, you think about reading the English. That's hard. But as ninth grader, they're good at reading English. They haven't quite acquired the syntax, the grammar, even the lexicon of what the asterisks means and what the x means. They haven't acquired the language of algebra just yet. And as a second language, it can make complete sense. So this inspired me to call this a kind of expert blind spot where one's expertise in the domain ability to solve that equation probably, you and I more efficiently, hopefully we would get them all right, but the equation might feel easier leads to our judging it as being easier for students. But the data tells us otherwise. Our intuition is informed and even in this case perhaps contradicted by the data. So expert blind spot, what I was calling it back then, but it's been a whole set of research on cognitive task analysis by Richard Clark especially. And he's articulate what had he calls the 70% rule which suggests that experts are actually unaware of most of what they know. They don't know about 70% of the knowledge that they have. The knowledge that allows them to perform in expert ways. In this particular example, we just saw, our expertise in equation solving involves a lot of things that our brain soaked up but we weren't aware of. That but to the extent that the 70% rule is correct, this makes teaching towards our expertise rather hard if we're only accessing the tip of the iceberg, the 30% we are aware of. And in turn, if data can help us dip under the iceberg to understand that other 70%, we have a chance to do better than with our designs. So what I want to transition to is talk about some online learning environments, the data they collect and also some larger scale evidence of this kind of data-driven adaptive support for learning by doing and being effective. This is a slide from a somewhat more project-oriented online activity in the algebra course that we created. And students are sometimes in more isolated practice problems that might involve solving an equation but this case, working on a bigger project where they're analyzing an authentic problem situation where they're asked to compare these two different cellphone plans. And they're working in these different representational tool, a table and a graphing tool that have different points in the course increasingly computer-like spread sheet capabilities and graphing calculator capabilities. More like paper earlier in the course. But as they're working through using the representations to understand the problem situation, they are getting feedback from this computer-based intelligent tutoring system that is tracking their progress in a way that I'll explain in a moment. but it allows the computer-based tutor to give immediate feedback to student students, for example, when the student forgot, has the per minute cost but doesn't have the $14.95 for the monthly cost. And when they rollover here, they get a feedback message. As they're working on challenging question, they might find they're needing some instruction and then the instruction then comes in the context of a need. If they're trying to figure this question out, they might be prompted to write an equation that sets the two expressions they wrote up above in the table together. And then use the equation-solving tool to solve that equation to find the intersection point. So that's one form of personalized instruction. Students get help, get instruction embedded by doing activities at the point that they have disrated need for it. And when they're cruising along, no instruction needed. There's an underlying model of the competencies, skills we sometimes call them knowledge components that we hope they acquire to demonstrate competency in problems like these and others. So those are two forms of adaptation that I'll describe in a moment how they work. First let me say, these data-driven design approaches have produced powerful learning results in a couple different kinds or paths of research. One has involved k-12 math tutors that we started building in the early '90s, spun off a company in 1998. And about a half million students a year over the past 10 or so years as it grew use that on a regular basis. And we did some early trials showing some positive effects but also a randomized trial was done a couple years ago. About 140 schools were randomly assigned to use the tutor or do algebra with the traditional approach. And the bottom line was result was that students a growth over the school year in the tutor course, and by the way, it's a blended course, it's not all online, just showed the online piece, but there's teacher professional development in there as a consumable text associated with it, students' growth from that course was more than double in the school year than a traditional algebra course. And another path of research has been developing online college courses and I know most of you are in higher ed. These are the open learning initiative courses here. There's 30-plus such course, many of them intro-college level. One particularly great study done by Marsha and others compared a traditional statistics course taught with about 100 hours of interaction over the span of a semester. As compared to the data-driven online course that was taught in a half semester without any extra time out of class which they verified in a kind of logging diary study. Yet, despite having only half the time, students showed greater learning gains. And the data was not only used in the design of the course, as I'll show in a moment, but also as feedback to the instructor before class meetings whereby the instructor could adapt what the discussion that day was about, similarly to what the adaptation we see with those skill bars where they now have a better sense for what students are struggling with and what -- [ no audio ]

>> Cognitive tutors use a cognitive model to adapt to individual student needs. And a model basically a computer simulation that solves problems in the various ways students can. And then in this example from equation solving, if a student reached this point for whatever reason, the system could project forward some possible next moves using a set of production rules that suggest those possible moves. So in green, I have two different correct strategies and in red is a typical incorrect move where the students forgot to distribute the 3 there to the minus 5. So if we have a model like this, there are two algorithms that use it. Following the students to the approach to the problem. When I showed you the complicated screen with the tables and graph, there's multiple ways students can solve those problems and the system follows different students down the paths. And when they need help, they get help associated with either the incorrect moves they're make, you need to multiply or in this case minus 5 by 3 also or a hint message, distribute 3 across the parenthesis and these get particularized into the contact that the student is currently in. So that's model tracing. The other is a outer loop that is assessing students' knowledge growth. So on each correct knowledge components here, we have an estimate for each student of where they are, how well do they know each of these different knowledge components? And that's used to individualize the pacing and selection of the activities and that showed up in the bar graph that you saw in the earlier screen. Hi does having this kind of data help us do better design? Well, what I want to show you first a learning curve. This is showing on the y axis, the error rate averaged across the set of students working through a curriculum. And in this cates, it's a geometry curriculum where they get success of opportunities to do steps and tasks. Not unlike what we saw a moment ago in the algebra interface. This doesn't look like much of a learning curve, does it? We would like to see the error rate going down as students get opportunities per task. What this is showing is if we thought of what students were learning as one grand geometry faculty, all one knowledge component and I'm showing a little slice of our data shop data repository here which you can get to I mentioned earlier in the up-front slide, if you think of this as learning geometry, the error rate on doing tasks goes all over the place. But there, I hope you're thinking, well, gee, the topics change in a unit in geometry area. So indeed instead of using just a single geometry knowledge component for which we get no smooth learning curve, instead decompose into topics or the underlying knowledge components that we think students need to acquire, then we get in the red here a declining error rate learning curve. Exactly the same data, just reorganized in terms of the different cognitive model or model of the underlying skills or knowledge components that students need acquire. And the upshot of this contrast is that if we use this kind of smooth learning criteria, we have a way to evaluate proposals for what are the underlying components of knowledge that make up one of these models? And we can compare alternatives to produce better models that both fit the data better but then inspire instructional redesigns. So you might be thinking, well, yeah, students learn differently but not just that. The real question is, what is a topic? What if what we think is a topic students find some tasks harder than others. Maybe that really means for the learner, there's more than one topic there and we need address those separately. Individual learning curves in this unit, some of them like circle area and trapezoid area, going down in error rate with opportunities to practice. But some like this one composed by addition are remaining flat. We have a problem there. So that's probably not on topic. So the goal for the instructor is to find out what's hard and redesign instruction. And just to give you the quick story on this one, the first time you're composing two formulas together is very hard. Repeatedly in the same problems, the next one are quite easy. If we scaffold the two areas you need combine in the interface before you combine them, that's of medium difficulty. So we end up splitting that quote one topic or one knowledge component into three. And then we redesign the instruction to focus more on the hard planning process and not worry about having students work on the easy stuff as much. And by doing so, shown in blue here in the treatment is increased practice on those composing a plan steps and in red is decreased practice on the simpler one formula execution, saving 25% of the time like graduating college in three years rather than four. That's a big possible reduction in time if this could be scoped out. And furthermore, the learning rate goes up. So we've done this a lot. We have a machine learning algorithm of searching through various models and what I wanted to show you next is a particular discovery that was machine-aided to get you a sense for the cutting-edge possibility for these kinds of tools. and I won't go into the details but I want to give you another poll here. The question we were addressing in the unit at some point was, do we need to present the processing of these area formulas for student, a separate topic to apply the formula in the typical way where you're finding the area? We would call that forward? Or verses applying where you're given the area and all but one of the other factors like a and w and you have to find the length. Are those worth separating as a topic? Or better to combine them? Because maybe after all it's the formula that you need to learn and you already know algebra. So what do folks think, should these be taught as a full group with perhaps separating each formula? Or splitting the forward and backward? Getting survey results, Padraig?

>> They're just coming in. So not quite yet. Right now, they're looking with the small end, well, let's see. There's sort of -- around 50% for rectangle, maybe a little more on forward and backward are equally hard for rectangle. For the circle, everybody thinks forward and backward are equally hard. 70 and 30 about. Trapezoid, backwards harder at about 57%.

>> I will say that when I was building this original unit, I thought that this would be a good review of algebra so I separated them. When we first ran these modeling tool, the data was coming back actually suggesting that they would be merged. But then when we ran this machine algorithm, we got a different result. It was suggesting that for backward circle area problems that they should be separated and we see that in the data because there is a big performance difference with students being about 80% correct. And this is averaging over their whole learning experience. So less correct in beginning, more at the end. But for the backward problem, only 54% correct. For other formulas, like a circumference problem, forward and backward were pretty even. So what we see here is a pattern where for many of the formula, that difference doesn't need to be made. But for backward, for circle, that is, circle area, going backwards the harder. If I give you the area of a circumstance and will ask for the radius, that's substantially harder. If you're interpreting this result, how do you interpret this result coming out of the machine-learning process in a way that's mean. And could help you redesign instruction? That's the key idea here. So do we have that survey up and going?

>> Nobody is biting on because circles are weird. 70, 75% because you need to take the square root. And about 14% because you need to divide by r.

>> Divide by r or divide by pie?

>> Sorry, divide by pie. Sorry. My screen is small. As a matter of fact divide by pie would suggest that circumference backwards would be hard. The square root would suggest that going doing the square area formula backward would be harder. But we didn't have such questions in the original data set. So provided a nice opportunity to see if this machine discovered data point and the interpretation which subject matter experts came up with. What we found is big differences for circle area forward verses backward but also as predicted for square forward verses backward. And pretty even for all the rests. So what do we do about that? We're going to redesign our system, the baseline, the control is our existing system that was doing this knowledge tracing for each shape as a single skill with those two merged. Actually the original baseline those two split in all cases. But we changed that on the early hand modified cognitive model. The redesign though makes a separation just for circumstance and will square. But we also made changes to the interface and to the hints so you see on the right, the new system has this in the interface some direction for how to go backwards with forward and backward. And the hints are more explicit about when you use the square root operation. And we did what we call an in vivo experiment in a local high school. Smallish but big enough to get interesting and generalizable results we hope. A pre-test random assignment and then a post-test. We call this closing the loop. The big important part of doing this data mining is to close the loop to bring it back to instructional design and show whether you get improvements with the new version of the system. And sure enough, pre/post gains were much bigger for the redesigned tutor than the controlled tutor when we're controlling for pretest differences. There's p less than .03 significant effect of the redesign. So it produces significantly better learning. From a machine discovered insight into the data. So I think I've eluded to some roles for instructors here but I wanted to illustrated how I've been using these high school math domains because we've more experience doing this and demonstrating how to close the loop. But we have data in data shop on in other courses at the college level. For example, there's an introductory psychology course. And here's a learning curve from that course where there is this, again, this blip upwards in error where originally the designers were thinking this is all one topic, describing psycho active drugs but the data seemed to suggest Otherwise. And the instructor's job is to figure out why are some of the questions much hard senator and which questions is it? And can we then see if some involve a different skill or concept? So there was a psychologist, went through the data and labeled some of the questions as involving a second kind of a skill. And I have to say, I don't know this particular content of drugs very well. So I'm not sure what the exact distinct is, but that's part of the point. You need subject-matter experts, it's a combination of data, machine interpretation, machine analysis and human interpretation. If you split those two as per the subject matter expert's reliabling of which tasks involve a different and harder skill, you now get two smooth learning curves. Really two topics if you will with underlying skills or knowledge components. So the data helps, you know, something like this, it can take years of observation and grading of homework assignments to derive some of these insights and the data can help speed up the process but the in strike or the domain expertise is super crucial to all of this. So with that, I hope you get a sense for at least in one of these domain, remember I indicated there are three different broad areas that we could do this kind of data mining and we have in various forms. But I want to hand it over to Padraig to say a bit more about making this work in practice.

>> Thanks, Ken.

>> Do that. I have to get out of here. Yep, go ahead.

>> And while I'm starting to share my screen, I just wanted to mention that there was a question in the chat that maybe you can talk about while I'm getting it up and. The question was, is this subject specific mathematics return Ands builds on time? Or are other disciplines the same?

>> Yeah, I hope my psychology example hoped to answer that question. We've done this sort of thing in statistics, in second language learn, in psychology, oh, in writing there's a writing course, particularly on prose-style tutor is part of it. And that English instructor little bit more technically oriented did this kind of learning curve analysis to identify for example how nominalizations make text unclear and how identifying nominalizations are easier in some cases than expose adapted her instruction to better address those hard cases so. Yeah, it's really is about digging in, having a good set of formative assessment activities and seeing what it's making good performance hard for students.

>> Great, thanks, Ken. So I'm going to zoom out a little bit and talk about quality learning. So a big part of what Ken has been talking about is, thinking about how to create these quality learning experiences through instruction using the insights from data mining to help. And I'm going to start just by taking a step back to talk about what quality learning is for engage. So we start with identifying where we want to go and then choose the best route to get there. And the best tools to help us. Obviously choosing the correct tools are very important. But yeah, so thinking about these topic, thinking about the learning objectives, that's the place to start. And we can see how using data mining can help us understand what are the appropriate topic, the appropriate learning objectives. At the same time, we believe that one learns by doing, quality learning is authentic. One doesn't study to become a pole vaulter by looking at diagrams of pole vaulting but actually practicing and getting feedback along the way. And we believe that learning solutions should be for everyone, which means we need to design accessible learning solutions. But it also means we need to design for all sorts of other differences. Students face all sorts of barriers. They are not equally prepared. Not all have their basic needs met. So how do we attend to their needs and we are taking a universal learning approach looking how to create multiple means of engagement, representation, action, and expression. And then learning solutions need to have elements that are tailored to the individual needs and interests of students. And Ken showed some of the ways in which the tools he's worked on are adaptive and have individualized feedback and that's a really important part of creating effective learning. So I'm going to take a deeper dive into this principle of personalized learning as an illustration of how some of what Ken discussed works in the context. Working on ways to make sure individual students get what they need from our learning solutions. Personalization is a response to the historical design of learning experiences in the U.S. which is emphasized standardization over everything else. And the problem is for in student, that doesn't work. Learners are not standardized. They come from a wide variety of backgrounds, experiences, abilities. So we need to come up with efforts to create learning experiences that are personalized and usually these days these efforts are focused on the power of technology to solve this problem. I think one of the nice ways that Ken approaches this is that he thinks about it in terms of how technology can help us solve a problem but not solve it all by itself. So the influential study that a lot of personalized learning advocates point to is bloom's two-sigma study that suggests that students learn better in a one to one learning environment. You can identify student struggle and mastery in the moment and intervene accordingly. Designing for how people learn means designing for how particular students learn. And we do this type of thing. We can analyze learner activity, provide feedback and recommendations targeted to that learner's particular needs. But as Ken described, this is predicated on our knowing something about how students learn particular things. And this is predicated on understanding the knowledge areas the topics of particular domains. Another nice example I have heard Ken use as well, consider how one learns how to ride a bike. What's fortunate know? You need to learn how to balance. You need to learn thousand peddle. You need to learn how to steer. You need to learn how to put on your helmet. So for the longest time, our [indiscernible] of how to learn suggested we learn steering and peddling first. And training wheels were a scaffold to take away the difficulty of balancing. Nowadays, we often see kids learn to balance and steer first and pedals come after and helmets come along the way, if at all. Helmets are interesting because they are a cultural aspect of riding a bike. So machine learning and analytics can help us understand different disciplines and domains and what's hard about them. But the truth is, we don't know all the answers right now. That's the point. We need to use these types of tools to speed up the process of understanding all of these domains. A lot of them are not really well formed. We don't necessarily know what is the right sequence of thing, what are all the topics? Our assesses are sometimes not as precise as they could be. The domains are not as well formed. Larry burger who is some of you may know, a CEO of amplify, talks about it being a GPS mapping exercise where the precision is bad enough that we might not know whether a student is in Maryland or Delaware. So there's a long ways to go. Machine learn canning help us figure it out. And can and some of us have talked about ways we can work together to extend what we know about how to do some of these things. But that's also incremental research work. And in the meantime, Cengage has a host of domains and thousands of students who need us to help them. Sorry, can everybody still hear me? I had a weird sound

>> Still coming through loud and clear.

>> Yes, we can.

>> Great, thank you. So we need a host of humans as well as tools to help us understand what to teach and when. Cengage has subject matter expert, learning engineers and involve faculty development partners in all stages of design. And help the us design, build, and test every stage. And we ask them questions about what data they need, what types of analysis, how and when it should be presented, what they're going to do with it. What else can we give you and how it can help. And while we knew it somewhere deep inside, something was confirmed through the process of designing and that's that the context matters. You know what's hard to teach in a one to one tutoring environment? Lots of things. So maybe one to one is not the only model. I know that personalized learning advocates often say, this is the thing that we're trying to do. We want to create one to one experiences for students. But we also know this, it's one to one is effective. We know it's not just about domain models and learning models and just in time individualized feedback. Maybe the students in bloom's study did better also because they felt like there was someone that cared about their success. Individual attention itself regardless of the feedback and instruction could be beneficial. So personalized learning is about activity in which you're treated as a person. And that's about competence in a domain or competence in terms of practice or knowledge. But it's also about being a part of a community and the relationships that that entails and it's about agency and motivation, the learner brings and the instructor can respond to us. For us, personalized learning is about increasing the number and quality of human interactions in which learners get feel back, and feel agency and feel connected. And this ultimately should be the purpose of all tools use for learn, tech any logical or otherwise. So I just want to say thank you. I think we have about 15 minutes left for questions and discussions.

>> Padraig, one question that came up with a lot of thumbs up was could you talk about mind tap.

>> Ky talk about mind tap?

>> Yeah.

>> A specific question about it?

>> Umm, let's see if somebody jumps in to try to clarify. It was phrased, could you talk about mind tap.

>> So mind tap is a one of Cengage's product which is a way for us to deliver our content. It's a tool for instructors and students. It's something that we spend a lot of time working on developing and improving. It's just one of the platforms we use. But yeah, it's one of our platforms.

>> Another question was, does mindtap have data that instructors can use?

>> Yes. We are using mindtap to collect data. I think it's depending on the particular course area and product as I mentioned, I think there's a wide variety of sort of definition and we built out and understand some domains better than others, but we are trying to increase the amount of feedback and data that we collect so that we can provide instructors with better information about what their students know and when and importantly, I think what we are really interested in doing is not just providing them with information but with providing them with tools that will help them do something with that information. A lot of things that we often see with online tools is lots of dash boards that give you tons of data but really what instructors need is data that along with something that they can act upon immediately.

>> So there was an earlier question, actually four questions about the machine learning algorithm. I guess to try to quickly highlight some of the things. The question about the cap value on the output and there's a sensitivity toward there about does the algorithm overfit the data, is it really capturing what wouldn't be present by chance? We use cross validation and other faster approximations to make sure we're not overfitting. Aver all, when you make a finer grain set of topics, you're going to get better fits to the data that you're fitting your algorithm, fitting your model to. But what we're looking at in cross validation to predict on held out data. So if those extra parameters over fit the training set, they won't do a good job on the predicted set. So the algorithm was asked whether it was a decision tree or lazy learner like ken. It's more of a search algorithm like in genetic algorithms that's posing possible cognitive model and is really driven by instructors are coding these questions as to what they think might be the relevant factors that are leading to difficulty or transfer of learn organize not. And then searching over all the possible factors. Certainly see some examples of that in some of the papers that I pointed to. So let's see, what else do we have? Does it have a learning analytics feature which instructors can use, I think was about mindtap, Padraig if you want to take that one.

>> I think that mindtap has some of that functionality. And we have another tool called learning objects which has a -- it's a platform that also has some tools for instructors that can help them in that way.

>> Yeah, looks like she has jumped in to point that out in the chat. There was a question about how much time is involved in setting up a course to use some of the data analytic tools? Well, I guess that depends on whether the course like the online course tools are already logging data. For example, in our open learning initiative courses, do have automatic logging of data. And then the question is, can it be transformed into a way that can be analyzed and part of what's in our learnssphere is the set of workflow authoring tools that involve data transformation. We've for instance, that psychology course was used as part of a mook. And we looked at whether students' different choices in using the learning resources available impacts on N predicting their learning outcomes. So for example, some students watch more of the lecture videos than others. Some read more of the online pages than others. So do more of the interactive formative assessments, the learning by doing activities with feedback. And we found those who do more of those learn by doing activities show six times bigger increase in their outcomes than the increase produced by or predicted by I should say, need to be careful, predicted by how much reading or lecture watching they do. Definitely is supportive of consistent with the idea that a lot of the learning is happening through practice. And that analytic workload is in learnsphere and could be replicated. Getting started whether the course ware provides that log data, a lot of it does. And doing transformations and getting them into this workflow. Yeah, I guess Joshua added to that, point accurately so. Great. So Aaron asked, speaking of one on one, I wish you could use this to chair dissertations. Yes, we're working on an intelligent tutoring system to advise Ph.D. students as we speak. Not really. Just kidding. But one thing I will say, that you know, the kinds of leadership and mentoring moves and activities that we do is dissertation advisers or instructors or any kind of more mentoring as said, a lot of that is involved with the personal in the sense of treating people as persons and addressing motivational needs and I think that's a rich area. We are learning a lot from social psychology research in that space on mindset, on bias issue, like stereo type threat, on sense of belonging, utility value, what you want me to learn important for my long-term outcomes? And understanding how we as mentorings make use of those things, you know, praise work rather than achievement is one of the mindset, growth mindset recommendations. There's a whole set of soft skills there that could be, I think not only more clearly articulated but we could be building more focus-deliberate practice activities for folks to become better mentors. I think it really comes down to what are some of the differences between really good mentors and others and how can we kind of capture those differences in ways that expertise can be transferred and. Not the by being told. Again, it's by having these kind of opportunities to practice and get feedback. Great. Well, are there any other questions? Or did we miss any in the chat there?

>> I think you've caught everything that's been submit sod far. But we have a couple more minutes so if any participants have any questions that they would like to share, feel free to type them into the chat box. And we have launched an evaluation link in the lower left-hand side of the screen, click there so you can launch the survey in the browser window and we would love your feedback on today's event. Just waiting a moment to see if any individuals typing have any questions or just thank yous. Excellent. Looks like we have reached the end here. So Ken and Padraig, do you have any closing notes you would like to share before we wrap up?

>> Oh, thanks for your participation. And attention. And hope to cross paths with some of you at some opportunity down the road. Oh, and on that score actually, we do have ways to engage with folks here through, for example, we have a learn lab summer school. A week-long set of mostly learn by doing, some general lectures on learning \science\signs and technology, but one of the four tracks is educational data mining. Actually, thinking of modifying that next year so we have a more practical learning analytics track in addition to somewhat more deep dive into educational data mining and machine learning and that will be up on the learnlab.org website. I think we just picked the dates for next summer but we don't have the application up yet. But that has a whole range of participants from faculty to grad students to industry folks and we welcome everyone who is interested. There's a intelligent tutoring authoring and online course authoring track as well.

>> And just add that it's been a pleasure presenting with Ken and thank you so much for your attention and your great questions. And please feel free to reach out to either one of us with follow-up questions. We would love to engage after this. And like Ken said, I hope we run across some of you elsewhere. This is an exciting topic, and there's a lot of new work to do and a lot of ways that we can see how quality learning can be improved and quality instruction can be improved using these methods. So thank you very much.

>> Excellent. And thank you to both of you for taking the time to speak with us today and sharing this information. It was really a wonderful session and I love to see the conversation that was unfolding there in the chat throughout as well. On behalf of EDUCAUSE, thank you all for joining us today for an engaging session and conversation. Before you sign off today, please click on the session evaluation link-which you will find in the bottom left corner of your screen. Your comments are very important to us. The session's recording and presentation slides will be posted to the website later today. Please feel free to share it with your colleagues. On behalf of EDUCAUSE, this is Adam La Faci, thanks for joining us today.

**End of Webinar**