Innovative Assessment of Collaboration 2014
Panel 2: Interactive Learning Environments

Alina von Davier

Thank you for coming back for our second panel today. Before I start introducing the panelists for this panel, I want to answer one of the questions that seem to be on almost everyone's mind. The question was whether the slides will be available and the short answer is yes. There is also a longer answer, but probably I will not go into the details for that for now. We will also have the video available. So we are building a webpage and we would want to make all these materials available. However we are also going to work on an edited volume on collaboration, and we will work with our panelists and perhaps with other experts that will agree to submit chapters to this volume. So we need to be a bit cautious about what is what it will cost versus what will be in the chapters. So the short answer is yes, and we will also have the videos and we will send you all, everyone who registered, definitely we have your email address and will keep you posted.

So all right, the second panel of this meeting will be focusing on interactive learning environments and I would add to that assessment environments as well, but we stay for now with learning because that's where most of the work has been conducted so far. The presenters, and I will introduce them probably more in the alphabetic order because it's easier for me to go by page by page. So we have Vincent Aleven, who is from Carnegie Mellon. Vincent Aleven is an associate professor in Carnegie Mellon's Human Computer Interaction Institute and he works together with, there she is, Nikol Rummel, who is a professor in Bochum, Germany. She's the head of Educational

Psychology Research Group at Ruhr University at Bochum in Germany. Dr. Nikol Rummel is a full professor and head of the Educational Psychology Research Group at the Institute of Educational Research at the Ruhr University at Bochum in Germany, and she is also an adjunct professor in the Human Computer Interaction Institute at the Carnegie Mellon University. Then we'll have Art Graesser, who is actually the first to speak in this panel, and I am sure probably everyone here knows Art. Art Graesser is a distinguished university professor from the University of Memphis and an honorary research fellow in the Center for Educational Assessment at the University of Oxford. Carolyn Rosé, Dr. Carolyn Rosé, Penstein Rosé is an associate professor of Language Technologies and Human Computer Interaction in the School of Computer Science at Carnegie Mellon University. And we have Saad Khan is a senior research scientist in the Center for Advanced Psychometrics at ETS within the Research and Development Division at Educational Testing Service in Princeton, New Jersey. So I will let Art start his presentation and we'll have the same structure of the panel as we did this morning in panel one. So each presenter will talk for about twelve minutes and then we'll have the questions for the panel, first for the panelists and then we'll open the discussion to the audience. Thank you.

Art Graesser

Can you hear me? Great. Our goal, we'll try to keep you awake after a great lunch and a great cookie. So what I'm going to talk about is collaboration through dialogues and trialogues with these agents. For those who don't know a conversational agent is like a talking head, or can be even an email message without a head. It's just some sort of communication by some sort of virtual agent. And for the last 17 or 18 years we've been doing this, starting with AutoTutor at University of Memphis. So this is our institute. It's very interdisciplinary and one point I want to make is you can only create these systems in an interdisciplinary village and you've got to collaborate. Otherwise you can't make these learning environments with agent-based systems. One partnership we have is with the Army. The Army Research Lab and University of Memphis has a partnership on building intelligent tutoring systems that scale up and we call it the Generalized Intelligent Framework for Tutoring, and if we can get all these people who build these intelligent tutoring systems to be on a common framework, the idea is we'll create better systems that will scale up.

So one reason they turned to Memphis is because of our research on agents, and we've built many of these systems, I think about 15 systems on a variety of topics. The first system was AutoTutor that held collaborative interactions between an agent and a student where they collaboratively co-construct an answer to a difficult question or a solution to a difficult problem. And this is a collaboration between two, namely an agent and a human where you go back and forth and back and forth, turn by turn by turn, maybe 100 turns, maybe 200 turns to answer one question. And since this is a learning environment what you do is try to get the human to actually articulate and do. So it's not just lecturing. It's trying to have a collaborative interaction. And we've done it on many topics: computer literacy, physics. My collaborator Vasily Russe(?) has a deep tutor in physics, biology. We have literacy and we have critical thinking. And a lot of these are really on topics that require reasoning, verbal reasoning. We're not so much interested in just presenting a multiple choice question. We want to have a collaborative interaction.

So here's just one example. This is a physics problem, a rear-end collision problem where a lot of people have misconceptions that when you get hit from behind your head goes through the windshield and what's the physics behind it? And many people have a billiard ball analogy where just like your hit in pool where the balls hit each other and your head goes forward, that's a misconception in these rear-end collision problems where actually your head goes back and it's the recoil that goes out. So to answer this question you have the main question where you have to explain your reasoning. You have your talking head. You then control parameters in a simulation like mass and speed and then you see the simulation, but you don't stop there. You describe what happens. We think that there's this inner mixture of activities where you're collaboratively trying to answer this question in the simulation. We think that's what it takes to get people learning. And once again we're interested in deeper learning. We're not so interested in shallow learning.

Now we of course assess what we do and this is one of my favorite studies – Carolyn Rosé was on it as well – where we compared four conditions in a randomized control trial with a human tutoring computer-mediated conversation. We had AutoTutor. We had reading a textbook for an equivalent amount of time versus doing nothing and this is what we find. We were happy to see that the human tutor and the AutoTutor were almost the same. But just as interesting you might have noticed reading the textbook and doing nothing. They were undisguisable. I mean shallow knowledge you get from reading textbooks and listening to lectures like this, but it's the deep learning that requires that interactivity and doing it with these agents is one way to do it. There may be other ways, but we wanted to demonstrate it in these conversations. And it's just not in physics. It's also in computer literacy, critical thinking. Time and time again if you want to learn at a deeper level it's not enough just to read or listen to a lecture. You have to have interactivity and agents are one way to do it.

So to do this you have to worry about speech acts, and one of the things we do since we try to understand the students' natural language is we have to classify all their turns into speech acts and assign them categories, and the computer has to do just something different if the student asks a question versus making an expressive evaluation versus a short response versus an assertion. It's a very context-sensitive, fine-grained interaction in natural language and you have to figure out through complicated systems, adaptive intelligence systems, how the tutor is supposed to respond, so it's a lot of fun. In developing these you have to have the agent give hints and prompts and pumps, all these speech-act categories to get the student to do the talking if you really want active learning.

So there's many functions of these agents. It can help when initiated if the user can click on something to get help, but one thing we know is many students don't ask for help under the right conditions so that's often not enough. A navigational guide, very often students don't know what to do next, especially in a complex simulation, so you want an agent to kind of nudge them on the next thing to do. We have modeling. You can have two agents interact and modeling good behavior, so they can learn from modeling. And then there's the adaptive intelligent dialogue and that's of course what we've been working with. And for all these systems we can have these agents take on many different roles: peers, tutors, mentors, bumblers, whatever you want them to do.

Now let me shift now from dialogue to trialogue because the recent years we've been interested primarily in trialogue. So we have the adaptive trialogues, and this is where you might make it vicarious learning where you get to observe two agents or you can have the human try to teach an agent. And in fact when the students are doing very well, that's when you want them to teach an agent. Those who can teach, so you can then have regular tutoring. So we have these rules that low ability you give them vicarious learning, all the way to high ability, the teachable agent, and now this is trialogues in learning. You can also have trialogues in assessment. So you can have this interaction, if we switch from learning to assessment, where you can assess how well they're doing, like if they give inaccurate responses, or take little initiative or violate social norms – that's low ability – all the way to high ability being the opposite.

Now we've been applying these in our Center for the Study of Adult Literacy, these trialogues, so you have, it turns out there's 90 million people who can't read at a level that gets them a good job. That's just in the United States. And so what we're trying to do in our center is to help people read at a deeper level. That's the Center, Study of Adult Literacy, and so you have a tutor agent, and you have a student agent, and you apply them all sorts of contexts, very practical contexts. This is reading commercial drug sort of descriptions to see if they know the purpose of the drug. And then you can even set up competitive game situations where the human is competing with a student agent and the tutor keeps track of points, make it very motivational. But there's these trialogues, and we can make these trialogues sensitive to emotions, not just cognition. So one of the things we do is track the emotions during learning based on multiple channels. I don't have time to talk about emotions. We're going to hear a little bit about this later in this session, but we try to make it emotionally sensitive as well as cognitively sensitive.

And one of the things with trialogues that you can do is set up these arguments between agents, so you can get a dispute, because very often deep learning occurs when people contradict each other and have a dispute, and to resolve the disagreement you have to have this conversation. And we've been doing research on this showing that something like confusion actually leads to deeper learning, and people, some people aren't, don't know enough to be confused, and so what you want to do is harness that for deeper learning.

So we're doing it in mathematics as well. But one thing I want to say, if I can get this clicker to work, is with Scott Paris' group we're applying the trialogues to assessment environments with ETS, and you can imagine a human and two agents doing along, and this is for English language learning where it can assess speaking, listening, reading and writing in this virtual world, and can they read that stuff. And here you have a doofus agent that makes a lot of mistakes, and then you have a smarty-pants agent that kind of saves him. And so you have this trialogue interaction as you wander around in a virtual world and from that you can assess reading, writing, speaking and listening.

So the last thing I want to do is just to say one of my roles has been chair of the expert group for collaborative problem solving for PISA 2015, and in there they're going to have agents, more in terms of chats, not the talking heads sort of agents, but more in terms of chat messages and emails, but they're agents. And the idea is to collaborate with one another and with two or sometimes three agents by holding this interaction, and by doing so, and this is the framework we have for collaborative problem solving. For those who don't know PISA is where you have these 15-year-olds in about 60 different countries trying to be assessed on collaborative problem solving skills, and this is going to happen in 2015. Educational Testing Service is creating the materials for this, and I've already seen a peek of the data that I'm not allowed to tell anybody about and we'll hear more about this in the next session. But as you can see, and I don't have time to talk about this, but there's two dimensions. There's a problem-solving dimension and there is a collaboration dimension, and those three columns, look at the core competencies in the collaboration dimension. I'd like to talk about it more, but it's time for me to sit down.

Carolyn Rosé

Thank you. I'm Carolyn Rosé from Carnegie Mellon University and I'm also president-elect of the International Society of the Learning Sciences. I'm going to start by talking a little bit about the context of the work that I'll be sharing about today, which is assessment of collaboration for the purpose of triggering dynamic support for collaborative learning. And in that discussion I'll talk about a three-dimensional framework that we have developed for being able to assess both cognitive and social dimensions of collaboration as they're reflected through the conversations that students have with one another and I'll close with one particular example that I think is a good one that shows the interplay between those cognitive and social processes.

So a lot of the work that we have been doing is in a classroom context in systems like what you see here where students are working together online, each with their own client interface. Groups of students logged into the same chat room, usually in pairs, sometimes larger groups than this, with a shared visual space and a chat space where the students can talk to each other and where there's also the presence of one of or more conversational agents. So in contrast to the work that Art talked about where he has multiple agents and a single human student, in our work we have multiple human students and usually a single agent or sometimes multiple agents that are there to offer support for the collaboration between the students.

Over the past seven or eight years we have done a series of such studies where what we've done is worked on design principles for the agents that support the collaboration. And I'm not going to get to talk about very much of this work in this short talk, but what I am going to focus on is what's kind of behind all that, what shows us how those conversations should look and helps to motivate the design of the agents by what we try to achieve, the difference that we want to make in those conversations.

So actually this work in computer-supported collaborative learning grows out of a longer-term engagement of me and my group in tutorial dialogue starting about 16 years ago, not long after Art started down that path as well, and in some of that work too it was important to think about what it is about conversational interactions that's beneficial for learning. And in those early days it was very attractive to think about very shallow things that were easy to count like how long were student turn links, how much were students saying in relation to how much the tutors were saying, how many questions were they asking, were they taking some initiative? And you can always make up stories for why these things should be beneficial, but what we found over a series of studies was that the effects, the relationships between these shallow easy-to-count factors and learning weren't consistent. So even though it's easy to make up a story for an individual study, it was harder and harder to make up a story that justified findings across studies. And that's what led to my group focusing on trying to find something that would be, that the goal was to find something more generalizable, that was more grounded in theoretical constructs both from the living sciences and from linguistics, and so I'm going to talk just briefly about this three-dimensional framework in this talk.

So it all starts out at the most basic level, thinking about two students within a group interacting with each other. What do we want to see happen there? And the most important thing that we want to see is knowledge building between these students. And so that would be the first dimension of this framework I'll talk about and we refer to that as transactivity, although the basic ideas you can find in many, many frameworks sometimes under different names. And then I'll talk about two dimensions from systemic functional linguistics that represent social things that happen within conversations that play a supportive role.

So the idea is that this notion of transactive knowledge exchange is the core of what we want to see from students. It's the main thing that we see consistently predicting learning across different collaborative learning settings. And here what we would like to see is that the social processes move students into a close enough proximity to one another in kind of a social space that they're willing to engage in this kind of behavior. Very simply it's about students displaying their reasoning to one another and building on one another's reasoning, and in so doing they examine their own thinking from a deeper perspective, but they also explore different ways that they might construct complex ideas together and then challenge one another.

So we do find evidence that this, that measuring this construct does have a generalizable predictive validity for predicting learning across studies, and we have done also a series of studies developing computational frameworks for being able to measure this in different kinds of collaborative discourse, including threaded asynchronous discussions, transcribed classroom discussions and speech from dyadic discussions, and I unfortunately won't have time to talk about that. But an important running theme through all of that computational work is how we borrow ideas from the underlying constructs that inform our computational models, and that it's not really about just having the fanciest machine learning model, but it's in how we represent the data informed by those constructs.

So in order to move students into a space where they're willing to engage in this behavior it does require them to feel a certain amount of safety and trust in their interaction. We start with what moves students along a horizontal dimension closer to one another, our framework from systemic functional linguistics called engagement. And the basic idea behind this is it's the kind of linguistic choices that people make when they present their ideas that show an awareness that what they're saying might not be agreeable to all of those present, and it displays something about their attitude about that potential disagreement, how open they are to the perspectives of others and these ways that people signal these things have been investigated extensively within that field. And we have seen strong correlations between our analyses of these language choices and how much reasoning students display in collaborative groups in correlational analyses, but we've also run intervention studies where we've manipulated the language choices of the conversational agents that are facilitating the discussions to see if it has a corresponding effect on how open students are and in fact it does, so we have evidence that it's a valuable construct.

There's also the vertical dimension and we call this vertical dimension in our framework authoritativeness. It comes from the systemic functional linguists framework called negotiation. I think I went the wrong direction. Okay. So again we have investigated this across a number of collaborative learning studies and have found how this measure of authoritativeness that shows how students are positioning themselves as sources or recipients within those encounters correlates with measures of self-efficacy and also makes predictions about learning, sometimes stronger predictions than self-efficacy does in those contexts. And we've also done a lot of computational work on this as well, not only in collaborative learning settings, but also in other settings like doctor-patient interactions and some simulations of work environments.

So I just want to very, very briefly talk about one example. I'm really not going to give very much detail at all about this example, but in this example we see how a manipulation that affects the social climate between students also affects the social positioning in those groups, and that that social positioning doesn't have a direct impact on learning, but it does affect how students engage in cognitive behaviors that do predict their learning.

And in this study what we had were students working on fraction problems together in a structured interface that comes from the Cognitive Tutor Authoring Tools. I believe Vincent will be mentioning that in his talk. But in this environment also students were interacting together through chat and what the experimental manipulation was there was simply that the agent in between problems would ask students in the experimental condition a couple of actually irrelevant questions from the math standpoint about their personal preferences which were then used to tailor the math problems and just make the environment feel a little bit more playful. And what ended up happening was this had a huge impact on how the students were willing to help each other. It significantly affected how much help they perceived that they exchanged within that environment. It also had a significant effect on how much help they actually exchanged. It had a huge impact on how much belligerence we saw being communicated between students in the groups. So in the control condition we saw students behaving much more competitively with one another, so the less capable student tended to end up being the, getting the brunt of a lot of sort of abusive behavior, whereas we hardly saw any of this in the experimental condition which was much more playful.

So we looked at the surface behaviors like the belligerence was affecting positioning within the groups and how that had an interplay with their response to coming to an impasse in their problem solving, and so from the conversational standpoint we coded the conversations with this authoritativeness coding scheme that I mentioned briefly. And this was a two-day study, and what we saw was that; so it was a two-day study and the belligerent behavior started already on the first day. And within the; in the control condition we saw a significant increase in aggressive behavior on day two. We also saw corresponding difference in balance of authoritativeness between our two conditions. So in the control condition we saw that the more capable peer adopted a significantly higher authoritativeness stance, whereas it was more equal in the experimental condition. And between day one and day two in the control condition there was a significant shift that accentuated that difference, whereas there was no significant shift whatsoever in the experimental condition.

We didn't see a significant impact directly of the authoritativeness of the student stance on how much they learned, but what we did see is that on day two when students in the control condition shifted down to a significantly-lower level of authoritativeness, with that how that ended up playing out for them was that when they would an impasse they would just abdicate to the other student. So they didn't get practice on just those skills that they really needed the practice on, and because of that students who were the less capable peers in the control condition learned significantly less than any of the other categories of students. And it makes sense that they would learn less if they got less practice on just those skills that they needed. And of course one could just stop there and say that the explanation for their learning was fully accounted for from a cognitive perspective in terms of how much practice that they were getting and it's well known that more practice leads to more learning, but if you think about the context and what was really going on there the reason for them getting less practice was a social reason. And so you can see how looking at the conversation from the social dimension allows you to explain a little bit more about what was going on there and also suggests a solution. It's not just enough to say, "Well, those students should get more practice," but you see how important it is to construct the environment in such a way that people feel safe enough to get the practice that they need.

So in conclusion I have talked a little bit about this actually rather simplistic three-dimensional framework for looking at collaborative discourse that we've been working on, and by no means do we feel that we're done with that work and we are continuing to elaborate it. But mostly what we're very interested in looking at going forward both in continuing in a classroom setting and in our current work in MOOCs is being able to explore this interplay between cognitive and social factors. And now I will turn it over to our next speaker.

Vincent Aleven

So good afternoon. I would like to take you on a little whirlwind tour of a project that my colleague Nikol Rummel and I are leading, and I should also acknowledge the other people on the project. And in our project we're studying fractions learning through collaboration, and the focus is on using multiple data sources to study how collaboration can be effective in this domain, and so I'll skip the big scientific questions and focus on assessment here.

So we have in one of our studies collected a dataset of small group collaborative learning, dyads of students working with an intelligent tutoring system, and we have data sources, like pretest and posttest data measuring how good students are at fractions problems. We have dialogue data. But in addition to this I would say more traditional sources of data as you would see in research on CSCL, on Computer-Supported Collaborative Learning we also have some other data sources including tutor logs, iTracking data and screen captures, and we're trying to understand how we can use these data sources and how we can take advantage of that to understand how and whether collaboration can be useful among these students, and what the learning outcomes are at the domain level and also how the two interrelate. That seems quite relevant to our goals for today.

And just to tell you how this dataset was collected, it was actually collected in schools, in two elementary schools to be exact, with 26 dyads of students in grades four and five who all took a pretest on fractions, then worked with the collaborative intelligent tutoring system for 45 minutes and then took a posttest. So that's how we got the data. And just to give you a sense, the system is an intelligent tutor for fractions. Usually this technology is applied for individual learning. Like individual tutors it has adaptive support for problem solving, so it prompts for steps, it gives lots of help with the steps of the problem to the collaborating students and it adds to that a form I would say of non-adaptive support for collaboration, so a collaboration script that helps structure the collaboration to some degree which is tied to the specific phases of problem solving. So that's where this dataset comes from.

And so let's dig into the analyses that we can do with these different sources of data and in particular taking advantage of the fact that we can look at the same problem using different sources. And let me start with tutor logs. So this is data that is traditionally used often in intelligent tutoring systems research, but not so much in CSCL research, and it enables us to see, to look for evidence of domain-level learning in the data from the tutoring. So this log data contains lots of information, all of the student interactions, all of the student responses, so attempts at solving, whether those attempts are correct or not entry quests(?) and so forth. And from the data we can extract using standard tools actually, like the DataShop, which is an online repository and lots of analysis tools. Using this DataShop, which is widely used in intelligent tutoring systems research and educational data mining research we can extract learning curves from this data. And again this is about domain-level learning. Our students are learning fractions knowledge as they are doing these collaborative activities. And for this it's necessary though to split out the interaction data according to the knowledge components. For instance we don't expect students to get better at fraction division when they're doing fraction equivalence. Maybe there is some transfer, but it's probably not complete.

So we label the data and the steps in their tutor problems with knowledge components, which is sort of the decomposition of the knowledge that they learn. And once we have that, and once we've met the steps of the tutor problems, students' knowledge components, we can actually look in the data at successive opportunities to apply each knowledge component to see if there is improvement as students get better, if the error rates go down for instance. So here you see an example of such a learning curve for identifying the greatest common factor of enumerator and the denominator, and as you can see over successive opportunities we see this downward slope which we like to see, evidence of learning of that particular piece of knowledge about fractions. And likewise if we average across all the knowledge components we also see this downward slope in our data. So here, and this is a standard analysis in the DataShop once you have made your knowledge component model.

So having these log data enables us to sort of ground our analyses of the collaborations in an understanding of the domain level learning that occurs as students are collaborating. It gives us a sense if students are learning. We could actually do a lot more with this data, with these learning curves and with the log data that we haven't done yet. For instance we could look at which knowledge components are most difficult, for which the tutor is most effective. We could also relate these learning curves, which are learning during the tutor work with what we see at the posttest. You expect to see some consistency, but if that doesn't occur that's informative too. And we also, as I'll show you next, is this log data can be really useful as we combine it with other data sources.

One of those data sources is dialogue data, so the conversation students are having. They're actually communicating over an audio connection, so they're not literally sitting next to each other. It's a network collaboration with headphones and such. And we coded the dialogue data according to the ICAP framework by Micki Chi. You may be familiar with framework. It's sort of a broad caterigorization of different types of learning activities according to whether they're passive, like maybe listening or reading, or maybe Art's textbook results is about passive learning. Then one level better you would say is active learning. The student does something, but it doesn't go much beyond learning the learning materials themselves, such as highlighting for instance. Then there are constructive activities in which students are beginning to construct meaning that goes beyond the learning materials, and at the highest level you might say is interactive learning where students build on each other's contributions to construct, co-construct new meaning. And as mentioned the higher the level the assumption is greater learning will occur if we are at the higher level, although it's not always borne out in practice.

And so what we did is take our dialogue data, code it according to these categories, actually slightly more fine-grained categories, and with this coding now we can answer questions like, “How frequent is each of these categories? Is fruitful collaborative talk correlated with higher learning gains on the fractions posttest? And does fruitful collaboration, fruitful collaborative talk co-occur with successful problem solving during the work with the tutor?” And I have time only to show you one analysis that actually takes advantage of three different data sources, mostly the dialogue data. But here we asked, “Does fruitful talk like these interactive dialogues, does that co-occur with successful problem solving?” And so what you see here is the number of errors that occurred with each of these types, and this is the number of errors that occur with interactive dialogue and you see it’s actually higher than for the other kinds of talk, so the answer is no. Fruitful talk, and particularly interactive talk, seems to occur when there are errors in the problem-solving process. Interesting finding illustrating that, well, first of all that fruitful collaboration may be occurring at these important points in the interactions where students make errors. Possibly this could mean that students sort of recover from errors together. And it also illustrates how the combination of log data and dialogue data can help us do analyses. We can relate what happens in the dialogue to what happens at the problem-solving level, and relate these two and thereby arrive at a better understanding of how the dialogue might and the collaboration might lead to learning outcomes.

So let’s move on to the iTracking data. So you may be familiar with this kind of representation of iTracking data, a heat map, so this shows the duration of the fixations in a tutor problem by one particular student, so the warmer colors means longer duration of fixations. This is just one problem. Here you see the heat map of different students working on the same problem. Looks kind of similar, although there is kind of this, you know, it looks like the student is not actually reading the text here, whereas this other student is doing that very carefully. Maybe the student knew about Art's research results, about reading textbooks. But now interestingly if I tell you that this actually is; I'm moving(?) data from collaborating students, so they were actually working on the same problem simultaneously. So another hypothesis might be that this represents some kind of division of labor that they had negotiated through their audio connection.

And now so one thing we want to explore here is whether there is evidence that gaze(?) convergence might be an indicator of effective collaborations. In other words the intuition, and this has been put forward by researchers like Newsley(?) and Yerman in Lausanne(?) for instance, the intuition is that effective collaborators often look at the same thing at the same time because they are sort of effectively coordinating their attention and that that's manifested in their visual attention. And so we can't really see that very well in these plots because this actually is data over a five-minute period or so. So over those five minute period, yes they looked at the same things, but did they simultaneously look at the same things? So gaze recurrence plots let us investigate that and it's a bit of an unusual representation. So let's say we have a student A and a student B, so we're going to give both their own time axis, so we actually have two time axis for student A, student B. And so in this space we have; so each point represents what student A was doing at a particular point in time, let's say T-1, and what student B was doing at a different point in time, T-2, so that's what this point represents. And if it's black, we'll make it black, if at that point in time the students were looking at the same thing, so if student B at time T-2 looked at the same thing that student A had looked at, at time T-1, so then it's black. And so this point doesn't say anything about what it was that they were looking at, just that it's the same.

And sometimes in these plots you see these sort of larger areas of black. Usually they're not solid black as you'll see. This means that over an extended period of time the students were looking at the same thing. So in other words useful gaze convergence occurs, and then if we draw the diagonal, then we can actually see this is where the time is the same for each student, so this is really synchronicity, and here we can see that there was sort of over, that in this particular time the students were literally looking at the same thing. And that gives us a quantitative measure of gaze convergence that we could use as a proxy or as an approximation of dialogue quality, and we could also overlay the events at the problem-solving level that we get from the tutor log data. So let's say these red lines are errors and the green lines are correct steps. We can see that here some dialogue occurred in between two errors and maybe that was useful dialogue because eventually they got it right. Now that's speculative, but it could be confirmed if we actually looked at the dialogue data.

Now the real thing looks more complicated as is often the case, so this is a real plot of the dialogue that we can study. One of the things we can do is now we can compare whether the high-performing students and the low-performing students differ in their gaze attention. So this is a high-performing guide(?) and a low-performing guide. You do see more black in this region, so suggesting that this is true, although overall that correlation was not confirmed. We could also look at different regions in these graphs to dig in a little deeper what might be happening so we could break it up into sort of regions of black and usually you get these regions. So if we just pick one we see that this is related to a particular sub goal within this tutor problem and we could look at why it is that we see sort of more black in this region and in that black, and now we're actually back to our students who didn't read.

So to conclude we see that this gaze convergence helps us to find time intervals with high and low joint visual attention. It's interesting to study the gaze data separately, but also to relate it with dialogue data. And our overall conclusions are that ITS technology can be a useful platform for CSCL research, so students actually did learn some fractions knowledge. And we think that having multiple data sources is useful both because each of those data sources is useful in its own right, and in particular the dialogue data and the tutor log data, but the combination is even more powerful as I hope I've illustrated. So maybe sometimes more really is more. Thank you.

Saad Khan

My name is Saad Khan. I'm at ETS. So I'm going to be talking about behavioral analytics and intelligent training systems. So it looks like we all agree that we've got to move beyond reading, writing and arithmetic skills and include collaboration and problem solving in the 21st century competencies. But the difficult part is how do you measure these? How do you measure collaboration or problem solving? These are difficult to measure in your traditional pencil-and paper kind of even adaptive tests because the kind of things that you want to measure in say collaboration would be, "What was the process that somebody took to get to an end goal rather than just an end state?" There are non-cognitive behaviors that are involved, things like motivation and self-controls. So if you're familiar with Carol Dweck's work at Stanford, she talks about the malleability of the human brain, and it is possible to inculcate things like the growth mindset. That happens when you reinforce process thinking, reinforce the fact that, “Hey, you took the right process to get to the end goal,” rather than, “You got the right answer.” To couple this with a lot of emotion and affective states because we can’t divorce that from what, emotions and affective states because people use that to calibrate their interactions in collaborative team settings.

So that’s a lot of challenges right there, but you can’t necessarily put a bunch of people in say a setting like a test and say, “One of the key elements about the outcome is going to be how many times do you have turn taking.” Right? So you can certainly game the system in that fashion. But what we want to do is go into naturalistic in vivo settings, like in your classroom, like student-teacher interactions, collaboration learning in team and workplaces. And then of course you’ve got situations like MOOCs and online learning and the flipped classrooms. So there’s a lot of data that’s coming through now which could potentially enable you to get to some of these processes and the non-cognitive elements of assessment. However the challenge is that it’s truly big data. You’ve got audio data, video data, you’ve got data coming through the simulation logs, and it’s really difficult to try and make sense of it. Some of the data that’s been captured, say pixel in the case of video data or just audio spectrum in audio, it’s very difficult to make connections between that raw data and the very high-level competencies that you’re trying to measure. In fact the raw data is going to be very noisy. So that’s in a bunch of slides sort of framing the challenges that exist now that we’ve tried to get into measuring these very complex and abstract concepts like collaboration.

And towards that end, the work that I've been doing has been to propose a hierarchical processing in analysis framework. So rather than making a direct connection between the raw multi-modal, low-level data, which could be once again video, 3D, audio and the like, and say communication, you have to try and build a structured framework which takes you from say the raw metadata to low-level features like what will be the interpretation outside of the video? [Words run together] …. …. it's different classifiers like for facial expressions. Same would be the case for say if you wanted to make an analog towards gestures, postures and speech which would then enable you to make some inferences about mid-level representations, the affective states of the individuals involved. Were people mirroring each other's behavior? Were they smiling with the other person was smiling? These are innate or subconscious signals that people use to calibrate their interactions. And then you have some hope of making inferences about somebody's communication competency.

Now the thing is people do this almost innately. So if you're familiar with Ryan Baker's work at Columbia, he would be using human raters to create annotations of coding of what looks like a great collaboration or people being engaged. Now human raters bring with them a history, a lifetime of how people actually behave. So they're using the …. instead of the context to make sense of what's going on. Recreating that in a computation model is extremely difficult and that's the kind of challenge that we've got over here.

So I'll give you a couple of examples of some of the work that I've done which follows this approach. Some of the work that came out last year at the HCI conference, and this is related with some military applications, specifically tactical skills. So how could you come up with automated analysis of performance of a bunch of say young military recruits as they engage in a military exercise? The kind of data that gets fed into a system like this would be your sensors like cameras, microphones, RFID to track individual participants' locations, which would then in turn be used to get to some level of behaviors or activities to say, "Okay, a bunch of people over here, the three people made a particular formation like a wedge formation, which is a technical maneuver, and how did they maintain that over the course of the exercise?" That's going to be driven or assessed based on a canonical execution of what an ideal exercise looks like.

So once again the data would be of the sense that you take video of the scene. Now the challenges are how can you isolate individuals from the background? So there's a lot of computer-vision work on that to be able to track individuals. And then if you have multiple viewpoints could you consolidate all of that information to get to a robust estimation of the locations? And then that becomes the raw metadata of the behavior analysis going from action detections like simple things like move to objective, run for cover, etcetera, to high-level things like a particular event, react to an IED on patrol, and finally evaluate it compared with a canonical execution.

In the interest of time I'll skip an example, but I think I covered most of that. And once you have all of that data processed you can go about and do an after-action review. So this is great for people who want to go back and say, "Sure, I have some automated assessment, but can I got back and tweak it and in fact maybe even make a more, high-level human judgments on top of that?" So you could take all of the data, recreate it in a virtual environment and review.

Now that was mostly tactical, but then I started to advance more into creating simulation training systems for social skills. Now this is where you want to bring in things like facial expressions, body gestures and the like. So can we take in HD videos of the faces to track and say the person was smiling or was looking confused or bored. Or could you take the Kinect to track articulated body posture and say what kind of gesticulations or gestures were being used? And could you take the audio spectrum to say what was not just said, but how it was said. The tone of speech is a great signal that people use to communicate. You could couple this with also wearable computing or wearable …. like some of the work that's come out of the media lab about honest signals and they use these integrated sensors. They call them the sociometric badges, and they're trying to pick up on things like enthusiasm and energy form the speech itself. So this is all non-verbal behavior. Instead of saying, "Okay, what was said?" sure, that's important too, but could you pick up on what was the sense of energy over there? You can also use accelerometers over there to get to the …. body orientation. So one of the constructs that I'm really interesting in collaborative group setting is how people are oriented towards each other. So we'd find, say you have three people in a group and you will use say your upper torso and the head to create a connection. So you're speaking with one person and your head is oriented towards that. So is your body torso. However if somebody else interrupts, a third person, and if it's not really an important interruption, you might momentarily just face the person. However your body or lower torso is oriented towards the original person. Now that's a social cue that I'm still engaged with this individual, but I'll attend to you and then come back. So those kind of things could be measured with say an accelerometer to figure out where your head is oriented, where your gaze is and where is the torso or the lower body oriented?

And then the challenge is could you take all of those different pieces and come up with a holistic assessment, a holistic picture of what was the person's affective or cognitive state? Towards that end last year my colleague and I, we published a paper at ICME in which we tried to get to multimodal affects. So could you take video and audio and get to a measurement of labels like activation, expectation, power and balance from those two modalities?

So I'll just briefly show a couple of slides on the technical approach and the customary equation. You've got audio spectrum over there which needs to be segmented to get to individual word references and then you extract a number of features like energy spectrum voicing. Now a word of caution. With multimodal analytics it's sort of a runaway train because you end up in a situation where you have huge dimensional feature vectors. This is quite different from your typical say item sort of assessments where you have maybe more controllable dimensionality of the space. Over here we're talking about thousands of dimensions, dimensionality …., so reduction of the dimensionality is a key thing. Same is the case with video. And then you have to decide on the kind of classifier that you would want to bring to bear. You could have a static classifier, like say …. vector machine, or it can be …., which is taking into account the dynamics of the data. And thirdly I worked for the dynamic because this is all once again process data, temporal data.

You do end up in a situation where you have to think about the fusion of the different modalities. You might have data coming in from audio. These are very different features from video features, so how do you combine them? Do you combine them at an early stage or a late stage? Each of them has pros and cons. Naturally our approach was a hybrid. We call it joint hidden condition random fields and the idea was to take feature data from each different modality and have a hidden layer graphical structure with …. …. variables …. They're trying to capture some mid-level representation. This is going back to my original talk about you want to create a graduated structure. So the raw features like the audio features or the video features are then integrated with potential functions which could give you some mid-level representations. And finally at the final label layer you are taking into account …. structures. And sure enough we got some significant improvement. I'll skip that and conclude with a couple of videos and disclaimer. This is work that I was recently involved at SRI before I joined ETS, but I'll still play it.

[VIDEO]

[Resumes while video is still playing.] …. …. that you can create if you have access to what the person is doing. So this is automated assessment of the real person is doing, which is …. …. [video still playing] so the avatars could put you or steer you down a path where you could exercise some of your social skills. And there are other examples that I'd rather skip, but I just wanted to highlight one of the things that Diego Luna Bazaldua, intern at ETS, had been involved in too. We're trying to study collaboration and influence of non-cognitive constructs, affective constructs towards the outcomes of collaboration. And I'll just conclude with a couple of remarks that this is an exciting time I think and the right time to study things like collaboration because some of the tools are just becoming available to measure the kind of constructs that I think people have been discussing over here. So with that, thank you very much.

AD: = Alina von Davier
VA: = Vincent Aleven
AG: = Art Graesser
CR: =Carolyn Rosé
SK: = Saad Khan
R: = Other Speakers

AD: So now we'll have a set of questions that will show up for the panelists. So here they are and they are in two places. I don't know if all of you can see them. So when we talk about collaborative problem solving are we talking about collaborative skills, assessment of cognitive processes or skills or learning? So that's one of the questions that our panelists came up with during our conversations before the conference. The other question is what are the differences between collaboration around assessment versus collaborative learning? What are the differences between human-to-human and human-to-agent collaborations? And what are the common technologies in collaborative learning and assessment? And I would like to open these questions for all of the panelists so please. You can also choose the question you want to answer or you can come back and respond to a question even if it looks like we moved on.

AG: I can start. Let me address two of them. The first one is for collaborative problem solving and what are we trying to get at? Let me just speak to the PISA 2015 assessment. We have to get at collaborative skills. That's critical in that. But also we want to do it in the context of problem solving. Some of our problems are more difficult, some of them are easier, but we want to make sure we get the collaboration in both.

Probably not as much learning. That’s probably; learning in the process is probably not what we’re trying to get at in PISA 2015. Let me hop to three though that has been somewhat controversial and that’s differences between human-to-human and human-and-agent collaborations. The PISA 2015 people wanted to use agents because for logistical reasons and timing reasons we had to get assessment with a couple of half-an-hour periods. So we wanted to put the people in different groups and different collaborative learning environments and do it in a short amount of time, so that’s why we used the agents. And one point I should say is if you’re a human paired with a bunch of other humans who are not very good, then you won’t be able to assess that human and that’s a challenge, whereas if you put him a diversity of context then you can get some hope of a true assessment. The other thing I want to say is there’s no other claim that humans are equal to agents. That’s never what the claim is. What the important question is, is to what extent you are covering the constructs in that four-by-three matrix so we can see how much collaboration between people, they cover those different cells in the matrix, and you can also see the coverage on the individual working with an agent, an individual with an agent to see how much those constructs are covered. So the question isn’t whether a human is equal to an agent – of course we know they aren’t – but more so on the coverage of the construct of collaborative problem-solving framework. So I’ll just leave it at that.

AD: Thank you. Nikol?

NR: I wanted to say something about the first question too, also reacting to what you said, Art. I think one thing that worries me a little bit about assessments of collaboration with the goal of just assessing basically the collaboration, the good collaboration as a goal is that a lot of the research on computer support collaborative learning has found that oftentimes when you try to support collaboration by for instance scripts or instructional means or training, you manage to improve students' collaborative learning. You manage to get them better as being good collaborators as we find in many studies, but then very often you also find that it doesn't correlate with their domain learning. And I'm wondering, well, isn't, and that also builds upon things that Eduardo Salas has said this morning, isn't what we're trying to measure here, also by including it in PISA, a means to an end? Don't we at the end want people to become better collaborators because we want them to be better team members who are then working towards collaboration as a means to an end? So I'm wondering what are we getting if we're kind of removing; well, you showed this matrix that has both, like the problem solving and the collaboration. And if we're trying to sort of decontextualize and measure competencies and collaboration, is it the impossible and what are we getting? What is the tradeoff?

AD: So Nikol is answering the question with a question, which I always like. So the other panelists?

CR: So I'm not sure if this directly addresses what Nikol has said, but I think there can be so many different answers to that first question depending upon what your goal is. And I acknowledge that there have been a lot of studies where it's been shown that it was easier to elevate the level of the collaborative processes and to show a significant impact on learning during that session. But I think that the question really is for longer term, beyond that one session, do they learn things in that session that make them better learners moving forward? It's harder to do that kind of research, but I think that it's likely that in the long term we need both kinds of skills. We need the cognitive skills. We need to be good collaborators. If we're good collaborators we can benefit more from working with another student also in a collaborative session. And it's also the case if you stop and have a really good discussion you might learn something deeply, but you maybe didn't have time to get to all the other things that you want to be able to get to. And so in the long term as we think about a student's trajectory through their learning, there's a very complicated time management problem there and a sequencing problem of what you should invest time in in the short term and what will ultimately lead to bigger improvements in the long term, and I think that we don't have good answers to those questions yet. And I guess when it gets back to the assessment question, I think that there's assessment for its own sake at a particular period of time, but it can also be important to assess whether your system is or manipulation is achieving a positive environment for learning now. And so maybe that's not so much a measurement as the skill of the learner as it is a measurement of how well the match between what you're offering and what the student needs at that time is going.

SK: I would just add to what Carolyn and Nikol said at this point that, well, I think perhaps what we're trying, we're talking about this measuring, all of the above, collaboration problem solving, collaborative skills, assessment of cognitive processes and skills, etcetera. I would add to that non-cognitive skills too, because I think things like persistence, self-control, engagement, those are really important skills that people bring when they are interacting with each other. And so in my opinion they really ought to be also amongst the kind of things that you want to be able to measure if you want to create a really comprehensive picture of what collaboration assessment will look like.

AG: Can I respond to one thing?

AD: Yeah, sure.

AG: Let me clarify that I certainly think learning in the long term consequences of an assessment are important. Ideally having assessment of collaborative problem-solving skill at one point in time, you would want to motivate changes in curricula where people would learn collaborative problem-solving skills, but the purpose of an assessment is really to assess where they are at some point in time. Now that being said, the question does raise whether we want to assess their learning of collaboration in that half an hour or an hour session where if they see other agents modeling good skill can they pick that up and adopt it? We're not trying to do that in this assessment for 2015, but in principle that could be something to look at in other forms of assessment. Can you learn how to be good in; assessing learning as opposed to some sort of construct at one point in time.

VA: I would; some of the things that Carolyn said resonated with me. One of them was that maybe collaborative learning isn't always the best option and that it would be useful for us to better understand when it is and when it is not. So there might indeed be situations where learning something by yourself is actually a better option, just more efficient let's say, or even more effective than collaborative learning. So I think there are lots of interesting open questions around not just what is good collaboration, but also what is collaboration good for. And another point that Carolyn made that also resonated with me is the notion that learning to collaborate might take a long time and our efforts at supporting that should therefore also be sustained and maybe in multiple contexts and multiple situations so students could actually get to see that maybe what's good collaboration in one context when we're learning fractions might not be so different from when we're learning something else. That seems like a huge challenge, but also an important one.

AD: Very good points. I was also thinking listening to the panelists that we've been talking about intelligent tutoring system, about AutoTutor, and we have the dialogue, so real classroom and real people, and then we had learning systems that use nonverbal clues for that. I was wondering looking at the last question on the screen what do you see as the common technology in collaborative learning and assessment nowadays that we are talking about iTracking and about multimodal analytics. So how do you imagine our next generation of experiments even and studies?

AG: I'll start and then you can give the real answer. Interestingly you can go a long way with something like multiple choice. And in the assessment for collaborative problem solving 2015 you can imagine a chat between a human and agents where at any one point in time they have a menu of options and they choose one of those options for chat or they choose one of those options for an action. You can go a long way in the clever designs that they've come up with at ETS in assessing collaboration in that way. So that's one way of doing it. A second way of doing it is what we've done with natural language interaction and the multichannel sort of emotion detection, but that's expensive. I mean you've got; it's expensive in terms of all of the, like we have a butt sensor. It looks how you sit in a chair. Sorry, body posture measurement system – that's what they call it – and there's the iTracking, and so that's expensive. A camera to get the face can be pretty cheap because a lot of places, a lot of computers have those, but it's a little more expensive doing it that way. Of course the neuroscience method, some of the exciting stuff, I was talking to Ron Stevens about some of his stuff. That is, and we heard about it earlier, that is exciting too. So and Steve Fiore mentioned it also. So there is a whole spectrum of these methods, but you don't want to dismiss kind of the constrained multiple choice kind of chat-based sort of way to go also because you can get a lot of information from that.

AD: Thank you. Nikol?

NR: So not at all in the spirit of dismissing anything, but to add to what you said I think that the reason for us doing the project that we're doing is that we're trying to use a technology that has been very successful both in diagnosing and assessing cognitive learning and in supporting it. That's what ITS systems do. They support learning, they support cognitive processes and they assess it at the same time. And basically the starting point for this project that Vincent was talking about was to try to build upon this and try to see whether it is possible to also make useful for measuring, assessing, diagnosing, collaborative learning as it occurs around tutoring systems and at the same time supporting the collaborative learning with them.

SK: I think Art you already pretty much covered the common technologies out there. The fact is these are not really common at all. There is not much research on collaborative assessment. There are tools like cameras and microphones that are being used in a number of different fields to get to behavioral assessments for a variety of applications, but I think what you see, the …. over here might be a good collection of people who really just started to use it for collaboration. There are other tools also. For instance one of the sociometric badges that I mentioned, that's another tool that people are starting to use to pick up on nonverbal communication. So a Kinect like a 3D sensor is another thing that's really handy if you want to get to posture, and gesture …. and the like. But to a degree the people who really want to collaborate are way ahead of the researchers who are trying to just assess collaboration. People use Skype all the time to, or Google Hangout at workplaces to try and collaborate on projects. People use in the computer science world …. software to code simultaneously at different parts of the world and they're able to collaborate even without actually really writing too much over email, not that they're really good at communicating over email. So there are a lot of tools that people use currently to collaborate and I think we as researchers are only just starting to utilize some of them.

CR: If I could just add. So I guess just to throw one more angle in I would say I think that there's an awful lot that you can learn and support through intelligent tutoring systems that students work with individually or even collaboratively. But I think that there's still value in looking more directly at the social interaction between and around students because that's a real part of the experience that they have in classrooms in reflects something about who they see themselves as learners, and what they're responding to and the choices that they're making about which behaviors they engage in or not that are beneficial for their learning. And for the sake of understanding those contexts and fitting into those contexts and reaching those learners I think we need to understand those things. That can be a separate issue from a specific assessment question that we're trying to address, but I think in the spirit of being kind of comprehensive in how we're thinking about it it's a kind of assessment that is important to be able to do.

VA: So now that the low-hanging fruit is gone regarding this question, so in terms of technology; I have actually two thoughts, but one is in terms of technology I think we have an awful lot of useful technologies. But if I were to single out one I think the sort of improvement in dialogue technology is, which I'm not an expert in, but my colleagues over here are, and I think that's going to be very important. Maybe combining technologies is another good avenue, so combining dialogue technologies with more traditional intelligent tutoring systems. But I'd almost say that the question isn't so much like what technologies, but what collaboration? Like what is that we need to support with those technologies and what are the hallmarks of fruitful collaboration? And that question sort of looks larger to me than what technologies do we need. Sol we can make a lot of things happen in dialogue and in collaboration, but what is it exactly? So where is pay dirt so to speak?

AD: Great. Actually I would stay a bit with the technology, though you already went to my next question. So what do you think about us moving towards a Holodeck?

AG: Towards a?

AD: Well, I'm supposed to be boss here and …. [cross-talk]

AG: I'm sorry, I didn't hear.

AD: Holodeck, the Holodeck.

AG: Oh, Holodeck.

VA: What is it?

CR: Maybe you should explain.

AD: Well, so some of you I've …. [laughter] So it's the Holodeck. It's the virtual environment in which you go in and you learn through a virtual reality. And I know it sounds bold, but that's my role here to provoke and have the conversation starting. But Joe here, who is attending our meeting, has almost a Holodeck at his fingertips, and he has an amazing simulator and we could definitely work with an augmented virtual reality with dialogue, with multimodal analytics, with traditional tests and eventually intelligent tutorial. So those things sound science fiction, but actually they are not that far from, time wise from where we are. What do you think?

AG: Well, I know that Scott Paris's group are taking big steps for that with their virtual reality and their trialogues. I don't know whether they're getting to the point of picking up and sensing what the human does. They probably are I don't know about it yet. But I think that may be an environment that will be expected in assessment ten years from now. Otherwise it'll look old fashioned, what we have. But let me ask you is ETS ready to put those out?

AD: Well, I'm building the team to get …. You wanted to comment on that, Carolyn?

CR: Yeah, I do want to make a comment about that. So I respect that work. I think it's great, really interesting work and I look forward to seeing where it goes. But in my own experience I think that if you contrast what expectations the students bring with them into the environment versus what does the environment actually provide in terms of sophistication of dialogue strategies or how immersive is the environment, I think the expectations have a much, much stronger role in what we've seen. So this is an old study. I've talked about it several times, but we did these comparative(?) studies where in one study we had agents that were not very sophisticated, but in one condition we told people they were talking to humans and in another condition we told them they were talking to computer agents. And the behavior of the students was completely different, and the students who were told they were talking to a human read so much intelligence into the simple behaviors that the agents did, whereas in the other condition they talked about how stupid the agents were, they played around with them and they just behaved completely different. Then in a second study what we did was we augmented the agents' behavior with a human so that we could make the agent much more sophisticated without actually implementing a more sophisticated agent, but in both conditions we told people they were talking to a computer agent and there was hardly any difference between conditions and how the students behaved.

And it seems so clear from that study that it was all about what the students expected, and I do think that students or people in general have an expectation of what kind of interaction they have with technology versus humans, and no matter what we build into a computer environment it's still a computer at the end of the day. And so the kinds of inter-subjectivity that a student might be interested in engaging in with another student to get to know their mind and get to know them is just going to be really different than when they're responding to a prompt.

And even just as a more recent confirmation of this, even a simpler agent we're deploying in an edX MOOC that's going on right now there's a Virtual Carolyn agent that facilitates some discussion, reflection discussions between pairs of students at the end of individual learning activities. And we told people that these were just agents, but some people must have just missed that and they just thought that Virtual Carolyn was actually me. And there were students who got into these chat rooms, and even if the other student didn't show up, they had a great interaction with Virtual Carolyn and were impressed with Virtual Carolyn. And I was really shocked that this happened because we told them. We even told them. But in a MOOC you can't get everyone's attention at the same time. People just miss things.

AD: This is also consistent with Eliza. That's quite old, the old system Eliza, and it's also consistent with a newer system developed at the University of Southern California where they build the medical, the nurse, which is a virtual nurse to help people with depression. And despite the fact that the people are told, even in some cases experts were told that that's a computer-generated nurse, then people still found that extremely helpful and were extremely open saying things that, well, surprised the researchers. Let's put it like that.

AG: Let me build on that because I just saw that and they recorded me there just last week. And the research indicates that people are more open and reveal more about themselves when they're talking to an agent than when they're talking to a human, think they're talking to a human, and that's been replicated in survey methodology as well. So this is a case where actually the agent may enjoy some of the advantages over talking to a human.

SK: And I wonder if that's because human beings expect that are just not sophisticated enough to make the kind of judgments that people make of others. So I think that's perfectly and a beautiful result that you've got. I just wanted to mention one thing which is that people use; if you're familiar with computer graphics there is a notion of the uncanny valley, which is that graphics look cool and believable up to a point where they become about too real, but not quite real enough, and that's where people get disengaged, and I think there's an analog to that for behavioral agents too. So the challenge really is how to cross that uncanny valley as we make more sophisticated agents. So if that's possible to really make things like; well, what's the term in agent-based systems? You can [can't?] scale things up. You can [can't?] have a human being do the kind of assessment or put you in exercises because that doesn't scale. And the other factor is you can't repeat the same thing. So you can have repetition of the same exercise for different people and controlling all the factors. So if I wanted to get to a sense of putting myself in a collaborative exercise, it will really depend on what my partners are, how I calibrate my performance over there. So it's really difficult in those kind of situations to disentangle the influence of the participants from my own collaborative competency. But if you could control that with virtual agents which are sophisticated enough to elicit appropriate behaviors, then I think that would be a great thing for agent-based systems.

AD: And now since we don't have a lot of time, I want to get back to the second part of Vincent's last comment which was to link it all back to the construct. And we had that question up for the first panel, that sometimes we get so concerned with a technology and methodology that we lose track of the construct. So do you have; what are your opinions about how are we addressing the things that our panelists from this morning discussed? How are we addressing them with the technology that we have available?

AG: I'll just start and mention that at least form the standpoint of PISA Collaborative Problem Solving 2015, the traditional sort of chat-based multiple-choice has gone the distance in my view of covering the constructs on collaborative problem solving that were articulated in the framework. So that's one way of doing it and there may be many others.

CR: So I just wanted to say I think an important thing to consider – this may be a precondition on the answer to your question – is that when you think about groups in a work context, you're talking about such a different animal than the kind of groups that we look at in the learning context in classrooms. And the learners are older. They have much different goals. They work together under very different situations and for longer periods of time. And also the kinds of tasks I think, and we didn't get a lot of insight into this from the short presentations – they just didn't get down to that level of detail – but the kinds of tasks I think that in the research on group work in the context that were represented by those panelists from the morning, they just value different things. So I think there are definitely some constructs that generalize between them, but I do think it's also important to respect the differences. And so sometimes I've heard from colleagues of mine who do more of their group research, look at the research we do in CSCL and say, "I just don't see anything in there that I'm interested in or that connects to what I do." And at first I felt bad thinking, "Oh, maybe we're not doing good work," but I realized eventually that it's just that we're really thinking about such different settings. So I do think that has to be a precondition on our answers.

SK: And just a quick one. What struck me was the way the first panel was almost unanimous that collaboration itself may not be a construct. It might be an umbrella term for a number of different constructs and behaviors. And I think that just fits quite well with how are we trying to go about measuring that, because it may not be that you can go directly from the kind of data that you can pick up in a multitude of sensors and, voila, you have great collaboration assessment, but you have to go piecemeal in a graduated fashion to get to, okay, maybe turn taking is one of the constructs that are predictive or evidence about collaboration.

AD: That's a perfect place to open in to the audience. You can comment. You can answer some of these questions or ask your own questions to the panelists. So at this point we are opening up for discussion. Joe, perfect.

R: Thank you, Alina. Yeah, I'm the Joe from the Holodeck, so thank you Alina. You can come see it anytime you like. You paid your federal taxes. It's yours. You might as well come see what you bought.

CR: We'll just beam over.

R: Yes, right. Saad, you brought up the idea of behavioral analytics measuring a number of things – cameras, mics, tracking, audio, visual, speech, motion, physiology. So the question for the panel is of all these things we could measure, if we had a group of learners trying to a task, which of these do we have data or would you speculate on would be the most predictive of optimal performance of all the things we could measure or do we even know? If we don't know, what would you speculate would be the most important analytics we could look at?

SD: Okay, I'll just take a stab at it first and the rest of the panel can correct me. I don't really know which particular modality is the key predictor of your ultimate outcome that you're trying to measure. In my experience it's really been a multitude of them. In fact the fun stuff is at the confluence of all of them. So if you want to say get to a sense of affective constructs which might be influencing collaboration or communication, you can't divorce affect that you're picking up from speech from say affect that you're picking up from body gestures or body language, or if you're trying to get to situations about measuring at the sim center triage abilities. So there are a number of different constructs that you would be interested in, things like communication and turn taking. However it's difficult to measure each one of them in isolation, in just say video. So you have to bring in the entire picture together and try and delay hard decisions or scoring as late as possible when you've collected evidence from all the different modalities.

AG: Let me say just something briefly, not about collaborative problem solving, what's the best, because I think it's too early. I agree it's too early to know. But let me just speak to learning-centered emotions, with some of my research with Sidney D'Mello. We compared the speech wave, facial expressions, language interaction and body posture and we found we got 96 percent of the way there by just natural language interaction and the face, the face at particular points for particular emotions like confusion, surprise and delight. Some emotions like boredom you don't get on the face. But body posture pretty much was redundant and the speech wave in that one was redundant also. But I think for collaboration I think it's way too early to tell.

VA: I'm relatively new to studying collaborative learning because I've mostly studied individual learning with intelligent tutoring systems technology. But as I'm moving into this field of collaborative learning, I actually like the fact that I can build on some of the things that the intelligent tutoring systems technology brings which is a fairly accurate and detailed assessment of learning of the domain knowledge. And I don't think this will be applicable in all domains, but in those in which it has been proven to be effective, like mathematics for instance, I find it provides a nice foundation for studying other or studying the aspects of collaboration that really make a difference, like using the data sources that I mentioned in the talk and actually many others have been mentioned. But to pick among those is actually really much harder to do, so I think it's useful to work with multiple data sources and I would expect that there will be some surprises. I mean it's sort of tempting to think that dialogue data is going to be where a lot of the action is to be found, but I think not exclusively and we might see other interesting things.

AD: Thank you. Paul?

R: Paul Gade, George Washington University and the Army Research Institute. You talked about affect and emotion during the process of problem solving and learning. What about motivation to do this again? Do people as a result of these experiments want to continue to do this sort of thing again? How does it affect their attitude toward the process of collaboration?

AG: Let me point out just one phenomenon that may have high relevance, and if Pierre Dillenbourg were here he'd probably have a lot to say. At the beginning of a collaboration you have a lot of good feeling introducing each other and small talk. Then you get to try to solving the problem and different points of view emerge, and sometimes that can be distressing, so a lot of, there's a negative affect. Now it can continue that way if you don't get along, and at that point you either go north and you get a good feeling on solving the problem, I guess kind of like a high, or it can go south where the aggravation continues. And so if people knew about some of these patterns of emotions, they may be conscious of it and take steps to fulfil the collaboration better. And I know Carolyn may have some data to indicate that learning, collaborative learning may be different than collaborative problem solving in terms of some of the emotion patterns.

CR: I think I'm not going to build on exactly that idea, but I was going to jump in with something else, which is I think that Art brings up a really, really important point about how it is that in many real and valuable collaborations in a learning context there can be these times that people get very frustrated with it and want to quit. But if they're forced to continue and are supported through that process it teaches them something that can be valuable. And actually in our capstone projects we have seen students go through this semester after semester after semester and have heard from employers that they value that before students to go through these processes. And I think that in terms of the kinds of the kinds of technologies that we've been talking about, I don't think we have to stop with the idea that when it gets tough people want to drop out because we can monitor for that trajectory that they're taking and try to identify those moments where more support is needed, and to try to scaffold and support the process to help them through that and that can be an important part of the research that we do.

NR: Yeah, and I think that's actually one of the reasons why I think trying to develop adaptive kinds of support for collaboration is also important. I mean that's where Pierre Dillenbourg has also done a lot of research because he has found that any time of good support which may be useful at the beginning when people are unexperienced collaborators, in the trajectory of learning to be collaborators you need to have adaptive ways of taking away that support. Otherwise they lose the motivation to continue collaborating on the task, so that's an extremely important issue.

SK: I think that's the key, adapting to difficulty. Like …. would say keeping people in a state of flow. So and I love the MOOC example because if you look at the dropout rate over there 95 percent plus and it's not because the material is not interesting. It's that individuals sometimes reach a point – well, most of them apparently – where the material is just too difficult for them or it's not keeping them engaged. So if there was a way to further engender moderation or persistence by adapting the content based on individual competencies, I think that would be, that might actually influence or revive the …. way.

AD: One question.

R: Well, given that affect has come in, I wanted to relate a couple of points that have been made about errors and conflict. So from Bob Biorke(?) talked about desirable difficulties in the facilitary role of errors in learning, but when we talk about collaboration in learning and collaboration in problem solving there's also the desirable difficulties that come from conflict. So conflict is a really rich concept and there's different kinds of conflict, so we want the intellectual and task-based conflict. We don't want the interpersonal kind of conflict that arises. So to the degree that we can differentiate that with any of these technologies that you're developing, all the better. Right now we have very good behaviorally-based measures that get at through self-report the different kinds of conflicts that emerge, that have been developed in the organizational sciences, but I have yet to see those integrated with the kinds of technologies that you've been developing.

AD: We agree. And I was just shown that we have zero minutes, so I guess this sounded more like a comment than a question, so I think please feel free to take on further questions with the panelists at our break while you are drinking the coffee or a tea. Thank you for your attention.

END OF PRESENTATION