[Auto-generated transcript. Edits may have been applied for clarity.]

Today, we have a very exciting panel to talk about a very important topic.

The topic is process as the construct explorations of AI-empowered assessment.

And our two panelists hail from Peking University in China.

Dr. Xiaoting Huang and, PhD candidate Yawen Cheng will be with us today to go over a series of questions that we think are

going to be really interesting for anyone who is just beginning to dive into the questions about AI,

AI-assisted learning, AI-assisted assessment as well. As usual

our session today will be moderated by myself and Dr. Carrie Holmberg.

And without further ado, and our audience, we're looking forward to getting us started on this journey.

Before I get started exactly into today's talk,

I always like to center us a little bit on some of the values and beliefs of our center and why we do the work we do.

We are deeply committed at the IAEP Center to deeper learning,

and that means we're interested in the notion of skills and knowledge that students in the 21st century must possess

to be able to engage in the kinds of jobs and civic life that are going to demand from them critical thinking skills,

problem solving skills, collaboration skills, communication skills, and probably most importantly,

how to take what they've learned in school and begin to apply it into the real world

and continue to be lifelong learners committed to the notions of deeper learning.

We always want to remember, though, that in any framework or any sort of values we have about what students should know and what they can do,

Assessment for Deeper Learning will play a critical role in prioritizing how we assess critical thinking, problem solving, collaboration,

communication skills, as well as the subject matter knowledge that we're interested in understanding how well our students are utilizing.

Of course, for us, it will always be formative in nature,

and we emphasize the idea of assessment that is a continuous improvement model focused on student work.

So that's kind of why we're here and what we do in our series.

It's my pleasure to introduce our panelists.

First I'm starting with Dr. Xiaoting Huang, Dr. Xiaoting

Huang is a tenured associate professor in the Graduate School of Education at Peking University.

Prior to joining PKU, she was Director of the Hong Kong Examinations and Assessment Authority.

Her research area spans a range of issues from test reliability and validity,

item response modeling, computational psychometrics to policy issues on the use of assessment data.

Thank you. Yawen Cheng. It's great to have you here,

Dr. Huang. Yawen Cheng is a PhD candidate in the Graduate School of Education at Peking University,

specializing in educational economics and management under the supervision of Professor Xiaoting Huang.

Her research focuses on creativity, measurement and development among college students,

educational assessment using large language models, and evaluation and analysis of educational policies.

At present, she attempts to develop a dynamic,

AI-powered assessment tool for creativity and exploring differences in creativity outcomes between human- human and human-AI

collaboration. So we are here to discuss and talk about and unpack

five questions and I'll introduce them now.

Firstly, how can assessment systems and digital learning environments move beyond measuring static knowledge

to support continuous process-oriented evaluation without compromising reliability and validity?

Second, what ethical and practical safeguards are necessary to protect student privacy in multimodal,

AI-driven assessment systems that collect physiological, behavioral, and emotional data?

Third: In what ways does the shift from assessment of learning to assessment as learning redefine the roles of teachers and students,

and what kinds of professional development are needed to support this transition?

Fourthly, how can emerging digital assessment frameworks ensure equity for learners with varying levels of access,

digital literacy or learning preferences? And finally, what lessons can be drawn from international models such as PISA's

digital assessment frameworks, and how might they inform national efforts to redesign assessment at scale?

Boy, Dr. Holmberg, these are not easy questions for anyone to answer, and I'm glad we've got the four of us here to work on these together.

So let's start off right away with this first question.

I'll read it again. And it's how can assessment systems and digital learning environments move beyond

measuring static knowledge to support continuous process-oriented evaluation,

but without compromising reliability and validity?

Take it away, our colleagues. So I guess I will take the liberty to try to

answer this question from my perspective and see if people can join me and join the discussion.

Actually, this is what our, my team is working on right now.

As Carrie just introduced Yawen, her dissertation is on how to assess creativity and using

in the context of human-human collaboration as well as human-AI collaboration.

So I think I myself experienced three generations of assessments. When I was a student,

all the exams I took were paper and pencil-based.

So the same items for everybody--it's very static.

The assessment itself, it's very rigid. And it's suitable for assessing curriculum-based knowledge.

But then digital learning provides the opportunity for

very individualized learning paces. People can learn on different paces and they can choose to learn whatever they are more interested in.

And even, some of the digital platforms, they can tailor make learning pathways for our students.

Some students may use more audio video materials, some students who like to read words better...

They are tailor made learning materials for every student.

So with that our assessment actually has become very individualized as well.

For instance, the computer adaptive testing. So we have a different set of items for each individual test taker.

And nowadays I think assessment or measurement is deeply combined with, deeply rooted in the learning process.

We don't even need to make items for our students by analyzing all the learning data on the platform,

that we store on the platform, learning platform. We can see how fast students progress, where they are having difficulties.

So process mining has become a new tool for educational assessment.

So nowadays we can look at not only how much our students learned,

but how fast they learn, in which, what kind of approaches they took

when they learned.

So first, and also for our students, instead of static knowledge,

they actually need to have a wider span of skills and knowledge in this digital world.

For instance, for reading, maybe ten years ago

students were given a passage and then after reading the passage, they answered a few reading comprehension questions.

But nowadays, students, they, in daily life, our reading tasks,

we need to browse through different websites or using AI to answer our questions.

Our reading materials have become so multimedia, multi-sourced.

So reading ability, reading literacy actually means first you need to know what to read.

You need to choose what to read. And then you need to deal with conflicts between different sources.

And later on you have to draw on information from different sources and

comprehensively understand, passages, graphs, or even audio video information.

So assessment sort of needs to capture these features.

That's why we analyze the learning process instead of giving them expert-made items.

So but the difficult part is to still keep high standards for our reliability and validity,

because right now, the process data we can capture and store and scale are still very limited.

We know response time, how many times students browsing a web page or the mouse clicks.

But then there are a lot of activities or behaviors that's offline.

So we are still not able to capture those behaviors.

So these digital online learning process data probably give us very

limited information on how students learn and how well they learn.

Yeah. Brent, maybe I should stop here and let other panelists join.

Absolutely. I'd like Yawen to have a chance to say a few things about this, and we are going to be keeping a lot of questions going forward.

So I'm going to hold my thoughts for now. Yawen? Thank you, Brent.

So, I have been thinking about this.

First I think, it is, increasingly important to notice that the dynamic process is more important and

meaningful and we have more things to capture and assess now then just the students' outcome.

But we have to take it more seriously, because the reliability and validity must be the baseline foundation.

So I think the first thing is that researchers have to be more careful and such as the PISA program they have made

solid theoretical foundation and such as the theoretical model of students test and the evidence model so as to have a more

solid theory and make different assessments more comparable.

And the second thing is that I think the new test, such as the process mining, we can make more connected

tests with the relationship of the new tests and the classic skills.

So, uh, we can, uh, be more sure of its reliability and validity.

So I think not only the researchers, but also the educational practitioners need to build more theoretical foundations

of the new tests, such as computer-based tests and processing assessments.

That's my little thought.

Thank you. That's a that's a very big thought, actually, and we really appreciate it.

In fact, to my knowledge, right now, the AERA, APA and CME standards for professional testing are being updated.

And there's some story that we're hearing

that there's going to be a fairly large portion devoted to AI in some of the new professional standards about thinking about issues of validity and

reliability. So more to come on that topic, but this is a great start.

Carrie, let's take the next question. What ethical and practical safeguards are necessary to protect student privacy in multimodal,

AI-driven assessment systems that collect physiological, behavioral, and emotional data.

Well, this is a very good question because right now this is sort of the gray area.

There are regulations, but still there are a lot of things we are not sure about.

I think IRB has at our universities, our IRB has put very strict

regulations on what sort of data we may collect from students, especially students of younger age.

If it's intrusive like brainwaves or heart rate, and also eye movement, those

equipments are actually intrusive and it's not very comfortable when you wear it for a while.

So, but besides that, whether the data is kept in a secure place, who should be able to use it?

Does it belong to the students themselves? Can they retrieve the data?

And also to what extent may our researchers explore those data?

I think these are the questions. We don't have an answer yet, but it's very important if we want to go a long way in that direction.

I think nowadays we can drive cars so smoothly, but that's because we have all these traffic laws, right?

So similarly, we already really need some laws or some authoritative regulations from the places like

AERA or the academic society to regulate what people do, especially those biological and emotional data.

Actually, it's very private. And it's possible researchers may, come

to a wrong conclusion and make mal use of those data.

So how to prevent those? I don't have a ready answer, but I think that's a very good question.

And the whole academia should explore this together.

You know, I'm thinking of right now back to those professional testing standards that

are being updated and that we will hopefully gather some guidance from

That the one area of validity evidence that generally isn't gathered for either traditional paper and pencil or computer adaptive,

or one could even now say process-oriented assessment data is consequential validity data.

So we go back to the old story between Jim Popham and Lorrie Shepard,

who in 1997 were writing about this question of what is the role of consequential validity in any sort of assessment system?

How would one collect evidence? How would one think about safeguards and threats?

So I feel like you're not alone with this, Xiaoting, in terms of no one really has the answer yet.

We don't quite know where the guidance is. I guess I am curious, just from Yawen's perspective, being kind of a young upcoming researcher,

how how are people sort of in your generation looking at this kind of data collection?

Do you think they might look at it differently than an older generation?

I'll say that I'm older and I might be more worried about this, but younger people are less worried.

What's your sense of how the test takers kind of under 30 would be experiencing this kind of

physiological, behavioral, emotional data collection?

Thank you so much. I think that we also

focus on this question because one of my friends, they use LAM, also large language models.

They use the analogy that they just hear our chatting and they then they know more about what we think

And also when I'm doing my experiments, I also hear about the participants in my experiment.

They have to agree to the experiment and I have to make sure that they are not that anxious

Thank God that they are in an experience other than in the real world.

I think this is also a very important and meaningful question.

And also with the development of the new technology.

We have also to make sure that the data collection and the usage have to

make sure that both the researcher and the participants are comfortable and

when you believe that our experiment and our research is not just exploration of the

private, but more exploration of the academic and the educational.

Can I add one thing? I think besides researchers and test takers, we probably need to discuss

keep the discussion open to teachers as well, because they are probably a major user of all this online learning data.

And if they use the data to label students or to punish students, which is even worse, I think that will be disastrous.

I've actually already witnessed some sort of mal use in that direction.

Teachers think, " Oh, no, I know better about how my students learn."

Yeah. So that's dangerous. I think that's a good point,

that if the models start to develop information systems that increase stereotyping or stereotype threat and don't

distinguish between the need for more information and the generalization that's being yielded to the teacher,

it could reinforce preexisting biases or beliefs.

And I think you're right, that would be a mal use. And fortunately, also, I wonder if it just wouldn't be construct irrelevant.

I mean, this is the other question I'm trying to figure out is I don't know what the constructs always are, but if the construct is reading ability,

I might want to know about the physiological, emotional, and behavioral non-cognitive factors that influence reading fluency.

But I would probably be as interested in whether or not they actually make a difference in reading,

based on an more objective measure than just how people feel.

So it's an interesting sort of trade off. But let's go on to the next question.

These are tough questions. And I think what's exciting is over the next several years, we'll all get to answer them together as we learn more.

But I really want to say to our audience,

I think that your focus on process-oriented data and thinking about assessment now as something that sits inside of the learning experience,

not outside of it, is a really exciting new idea.

Let's go to the next question. Just a minute. All right.

In what ways does the shift from assessment of learning to assessment as learning redefine the roles of teachers and students,

and what kinds of professional development are needed to support this transition?

Well, I'll speak from the Chinese perspective, because in China, teachers are the authority and teachers are supposed to be very knowledgeable.

Teach us what to do and tell us what's right and what's wrong.

So that's how I grew up, and probably that's how Yawen grew up as well.

So nowadays, we have all these digital teachers who can give us very individualized instruction, give us very individualized practice

exercises.

The real human teacher actually does not need to be so knowledgeable or authoritative.

The teacher's role as far as I understand, is now a companion,

who learns with us, who helps us to focus, who helps us to find the right direction,

reminds us what to pick up and focus more on.

So basically, teachers are becoming a companion, a friend, a knowledgeable guide, instead of a very authoritative expert.

So, teacher education.

Actually, I think teacher education is picking up this message, and they are trying to

put a lot of emphasis on how teachers may make best use of all these digital learning platforms,

but then maybe from more philosophical ways, how teachers view themselves,

how teachers view their relationship with students as well as parents.

It's not formalized in a course in teacher education, but that's what we need to adapt in the next few years.

I think that's very important. Maybe it's a different story in US, Brent.

You're on mute. Do you have any thoughts about this topic, Dr. Holmberg?

Yes. I want to focus on the second half of the question.

What kinds of professional development are needed to support this transition?

The track record, so to speak, of professional development of in-service teachers has been mixed,

even though there is a research consensus that prolonged,

ongoing months of repeat kind of workshops and learningis what does it for teachers.

That's still not the norm.So on the one hand, I am a little bit down about like what

can professional development as it really happens in the US for in-service teachers really improve.

On the other hand, there's never been a better time for can that professional development being individualized by AI.

Can the AI assisted tools like be part of that transition?

If we can figure that out, which is a big ask, that can make a huge difference.

But it's anyone's guess, I think, how it'll turn out.

Can I ask a question? I'm very curious about how us teachers, like elementary or middle school teachers,

to what extent are they using AI in their classroom instruction?

Well, I think, go ahead, Brant. I was going to say, you can answer that, Carrie

Go ahead. Yeah, well, I can answer that and just say that Brent and I have been recently

interviewing teachers at the secondary level, so I can't speak to the elementary.

And at the secondary level, it's not as prevalent as as I would have guessed in terms of actual use,

that teachers and students together are using AI tools.

Although students are of course, using it on their own outside of class contact hours

Ubiquitously. It's common.

Because at Peking University, our students are using AI much more extensively than their instructors.

Yes. And I think I think we need to differentiate again what's going on in colleges from what's going on in the K through 12 setting.

So I think we all know that college students develop mentally with many capacities,

are probably more AI forward in their thinking and their uses and

their engagement.

But when you study what's going on in middle and high schools,

which we're attempting to do for our next book, we're learning that a lot across California.

Cross-sectional samples urban school, rural school, suburban school, school located near Silicon Valley,

school not located near Silicon Valley. That to a tee,

most of the teachers are still not reporting to us that they are doing much more than occasional planning with AI companions,

but they're not engaged in instructional delivery either co-delivery or not, or even moderated delivery.

When you mention AI companion that students are using in China

I'm wondering how far this goes down the system. I could understand in college they might be using such tutor bots,

but are they doing that in middle schools, for example, Xiaoting? Because they're not doing that here per se

Commonly. Not commonly, an exception would be Khanmigo. There are a few districts...

The depth of penetration is questionable. So just curious about China.

Yeah, I think in China it's also only a few schools or school districts that are using AI tools extensively,

and usually they collaborate with a university like the Beijing Normal University, Brent, you visited that university, right?

So as more like a research site or experiment site, but for the others,

for the majority of their schools at K-12 levels they use digital learning platforms.

The teachers use that to plan for their lessons, looking for different resources or instructional materials and

and that, I think that's the majority of teachers, what they do right now.

There are a lot of universities and a IT companies that are working on

developing those digital teacher or educational

large model, large language models, educational local models.

So I think we will witness huge developments like very drastic developments in the next few years.

You know, it's an important question not only for us as researchers who do work in assessment, but also for policymakers.

So one thing we have to think about, as we know from our training at UC Berkeley,

is that there's always a policy environment that's at sometimes the school level,

the district level, the state level, it could be at the provincial level and China could be at the national level.

There's many different levels. And I'm just curious, you know, what are you all seeing on this question in the sense of a policy?

We're seeing a lot of holding off.

There seems to be a kind of a certain amount of concern about going too fast, too quickly with adoption of AI-assisted learning.

And I say that there are exceptions, like you mentioned,

where a university like Stanford partners with a particular charter school, those exceptions exist.

But if you look across most of the systems, you see a reticence, a concern about policy.

So what's the policy environment? We'll move on to the next question.

We don't have to answer this in depth. Maybe we just raise this question. Is it well-defined in China yet?

I think, national wide, we kind of embrace this AI era, and our national policy is to putting a lot of resources in

developing AI tools for our educational system throughout all elementary and secondary and tertiary sections.

So, um, I think one reason, for this is because China, we are--you've traveled across China, right?

So our provinces are different areas.

They, develop at a very different speeds

And AI systems help the educational equity.

So for those poor rural areas of the mountain areas, they actually need all these resources to help their students learn.

So I think that's one of the major reasons why we spend a lot of resources.

We are going to spend a lot of money in this area.

Yes. I was fortunate enough to actually be in Yunnan Province and work with their Ministry of Education on a consulting project,

and I was amazed at how far ahead they were in a lot of areas, what I would call vocational or new vocational learning, applied learning.

And I suspect that they're going to leap ahead as well as your western provinces will.

But I also was noticing at the same time that while there is a certain amount of

shift going on, there's also a lot of static going on around things kind of being the way they used to.

We'll leave that open for the future.

As a question of research, let's move on to the next question for now, though and continue this interesting set of topics.

Carrie?

How can emerging digital assessment frameworks ensure equity for learners with varying levels of access, digital literacy or learning preferences?

Wow. That's very challenging as well.

All questions are very big and difficult questions today.

I think just as with our human teachers and traditional assessments, we try to meet the needs of different learners.

So, for instance, when I worked in Hong Kong Examinations authority,

we make larger prints for students with visual impairments.

And also we allow more time, examination time, for students with reading difficulties.

So to tailor for their needs.

But then digital platforms or the AI-assisted assessment systems should allow us to do more.

But then is it fair? How to ensure the equity for learners?

I think that's the difficult part.

Because I actually observed some abusement, like students who apply for reading difficulty.

Actually more and more students apply for that for more time in the examinations each year.

So we see a growing number in that.

So but who should make the decision on how to ensure that it's fair to everybody?

I think that's the the difficult part.

And so I don't have a ready answer for that yet.

Well, I really appreciate the starting answer because it's an area that we're exploring right now in our research on accessibility.

One of the important claims that's being made for AI is that it increases almost exponentially the potential for individualized accommodation,

for students with identified learning disabilities and also possibly students who are adjacent to those labels and identifications.

You know, not everyone who has a learning disability is diagnosed. In fact, not everybody gets services.

So this might be a new frontier of providing more, for example, diagnostic work on

dyslexia so that while a student is taking the examination or is engaged in embedded process oriented assessment,

it turns out that there is a concern the AI can flag for us.

"Hey, there may be an underlying dyslexia issue that's going on" that wasn't even diagnosed originally with this learner.

So there are great frontiers, great possibilities of improving how we support those who have not traditionally been supported.

But, Xiaoting, I think you're really right, because once you start equalizing those playing fields,

you also might be overstepping and giving more support to somebody who perhaps

shouldn't have necessarily gotten that label or that identification or that particular,

advantage. We used to think in terms of simple accommodations.

Around time you get extra time if you have a diagnosed learning disability to take a reading test.

But when you talk about process-oriented assessment, what you're researching, you start to wonder, like what will be the possibilities,

as well as the pitfalls of real time interventions where you are assisting that person,

but the machine is doing the assisting and should the machine be doing that, assisting

if we're trying to evaluate the so-called real, authentic learning of the learner?

So, I don't know, I just threw out some ideas there. Anything make sense what I said?

Yeah. But I think that really depends on how we build the models.

Right? So it's very

early and very difficult for us to make a judgment now.

Exactly. Yeah. And I think Brent, what you bring up goes to the core of the question of the purpose of assessment.

Right? And how stakeholders and people involved, they naturally have different perspectives on--even if it's an agreed upon purpose--

even if all those groups of people can come to an agreement, they still, don't have the same perspective on it.

And I think that's one of the tensions inherent in public assessment that involves large numbers of students.

Yeah, Carrie, I agree with you. That really depends on the purpose of assessment as well.

So, nowadays we can see, high stakes, large scale standardized testing.

Actually not as important as it was ten years ago.

So maybe later on we don't need so many formal assessments.

But if the assessment is combined or interlinked with our learning process, then probably we're not so concerned about fairness.

I think that's a really important point, Carrie.

And Xiaoting, particularly when you think about purposes or whom?

So if I'm the government and I want to sanction in the national exam the Gao Kao that I believe everyone took this test fairly and appropriately,

then I might have a very different policy towards AI-assisted learning than if I'm in a local school that's very interested in project-based learning,

and I want to make sure that somewhere in Shandong Province,

these kids get a chance to develop a science experiment using all the possible AI tools they could find.

The classroom and school purpose

for assessment would be very different than the national examination system or the international system that PISA and other batteries try to

capture what nations are doing. Um, So, yeah, I think purposes are going to matter even more.

It's an interesting question. Well.

Speaking of PISA. Carrie? Okay.

What lessons can be drawn from international models such as PISA's digital assessment frameworks,

and how might they inform national efforts to redesign assessment at scale?

Well, I use PISA data for my dissertation, which was a decade ago.

But, PISA is actually

what we look for when we have a question about what sort of new assessment technology are people using across the board?

So, I think at least for researchers we look for PISA's definitions of those new constructs like collaborative problem solving,

creative thinking, and learning in the digital world for this new restaurant.

So these give us a lot of insights and inspirations what we might do with our assessment technology development

as well as how we redefine what are the core abilities and skills or competencies our students may need in the future.

So I think theoretically PISA gives us really good

PISA is a very authoritative and very good reference for researchers and educators in our country.

And I think that's more for the academia and for research,

I think for those high stakes exams, we sort of want this to be stable.

We don't want like huge changes for our students.

So I haven't seen any redesign in our high stakes, national-wide exams like Gao Kao that use any

PISA's technology or digital assessment framework.

So I think these are these also related to the purpose of exams?

I think it's the direction, but it will take some time for us to really apply what PISA is suggesting in our

standardized or high stakes exams.

You know, I might look at LinkedIn too much, I probably do.

I hate to admit that's a platform that a lot of professional solution providers have been utilizing almost...

I mean, it's the daily updates are hard to keep up with.

And one of the things that came through that was, that media was, the new OECD AI literacy frameworks,

and I started to feel like it's a bit of a race to who can come up with the biggest, most comprehensive, most promising AI literacy framework.

And certainly that raises a couple of questions about how authoritative those bodies are.

For whom? For the US, I don't think that we typically adopt OECD guidance, although we may, because some of the researchers,

apparently are from some of our own top research universities and they've contributed

what do we think, just even about the idea of these AI literacy frameworks

how are they going to steer us as researchers, and your own experience towards what to pay attention to?

And I would opine a little bit that they may also take us the wrong direction.

And I'll say a little bit more about that, but I want to hear your thoughts first. Like,

do you do you think that we need to find the one perfect good AI literacy

framework that somebody says in the international community is the right one?

Or do we just sample from all the frameworks we can find and create our own national and regional solutions?

What are you thinking about that? What's your thinking? I think we probably won't be able to reach a consensus on that.

Yeah. So first of all, it really depends on that country's level of usage and their national wide policies

how to what extent to use AI in their education.

And I think it also relates to how much attention they pay to PISA exams.

I know in the U.S. student test taking motivation is a big issue.

So a lot of the students, they just don't take PISA seriously.

But here in China, we take the exam very seriously.

And so it's sort of a competition for our students.

That's why the data we get from PISA is very reliable.

So that's why we look at PISA's analysis--how they analyze the data and their analysis tools

how we may use their analysis tools to do the same thing to our national education monitoring systems as well.

But as I said it, we really we probably won't be able to reach a consensus on what sort of AI literacy or digital literacy is

depending on how much we want to use it in our classrooms.

Right. And also we won't be able to reach a consensus on whether PISA's

analytical tools are the best one or the only correct one as well, depending on the data structure, what sort of inferences we want to make.

So yeah, I think that's a really open area for discussion and what sort of things we, we want to pay attention to.

I've always tended to believe that whoever controls the test controls the standards in the frameworks rather than the other way around.

And one would hope that there's alignment between the items and tasks and activities in

any test regime and something called the frameworks or something called the standards.

But my concern, Xiaoting, which wouldn't surprise you given where we both got our doctoral degrees from,

is on the construct validity for many of the arguments being made on behalf of assessment literacy.

And one example of that is I find that in most of the frameworks I've seen so far,

including the digital assessment frameworks or the digital literacy frames, there is little attention on the idea of learning progressions.

There was little attention on the idea of developmental levels

where a beginner who is interacting with AI emerges over time into more expert practice,

even if they're a young person who's working through the middle and high school system and then coming to college.

There seems to be little sense of the developmental in many of these things,

but rather these seem to be even if they're open ended items and even if they're scored polytomously

they don't give you the feeling that they're thinking too hard about how a young

person acquires literacy in multiple domains across multiple contexts over time.

And so I worry about the pitfall of racing towards another set of items that don't

actually capture the learner and his or her development as a digital citizen,

or as a digital researcher, or as a digital learner. What's your thought about that?

You're shaking your head, and I hope you're knowing where I'm coming from with this.

I agree with you. But I think they tried in the 2025 learner in the digital world assessment framework,

they claim that they want to assess students learning or progression.

So they were given a new platform, a new task.

The first part is for students to acquire new knowledge.

The how successful this attempt is is still a question.

They haven't released any data or results yet. But given the very complex framework and limited time, I think that's a nice try.

We will see. Yeah. But I think

it's very difficult for us to develop such test items or tasks and

also the analytical tools

we need to analyze students' developmental trajectory is actually very difficult as well.

So, for instance my team is trying to use Bayesian models to do this kind of analysis,

how students collaborate and then their later efforts is actually related or conditioned on their earlier achievements.

So, but the analytical tool is very complicated.

Yeah, yeah. I mean, the standard methodologies that I was trained on, IRT analyzes multi-dimensional IRT analyzes item calibrations.

Well you know, the task calibration, person proficiency estimates, you start to wonder.

You're shaking your head again, like,

whether these are things of the 20th century and that they're not going to make it into the 21st century as ways of again,

coming to judgment about validity or reliability or even bias in items.

So, you know, like so just tell me more. What are you thinking about on this topic?

I think these traditional statistical tools or traditional construct theories are still very useful right now.

We need both the top down approach from those traditional theoretical based, construct theory based,

approach using the traditional IRT modeling to ensure our reliability and validity.

And then the other way, the data-driven approach is the bottom up approach

comes next to make our story fuller, to make us to make the researchers get a fuller picture about how students learn.

So I think both are important right now, but we cannot rely on only one approach.

Yeah. Yeah, that makes the whole thing difficult.

It's really interesting because I do think that we're looking at these sort of almost tectonic shifts in paradigms.

I believe that this is a paradigm shift. That's a very important concept.

Going back to Thomas Kuhn's work on the structure of scientific revolutions.

The data revolution is actually upon us, and we're using concepts like top down,

bottom up, or we're using ideas like, you know, qualitatively situated

versus sort of like quantitative.

If anyone is interested in asking questions or thinking about this with our colleagues.

Now is the time. And maybe we can also, while we're waiting for those questions, summarize.

So, tell us where you think we are, and where we ought to be next.

As our guests, we're so grateful to have, you coming from Peking University to share ideas.

I know Yawen is back now. And so if anyone has any thoughts or concluding remarks, let us know.

What's exciting to you about the future of this work that you're doing?

I think we are going to see more applications of AI in education and we should prepare for that

research wise, as well as philosophically.

So I think what we are doing is very interesting and meaningful and worthwhile.

So, Yawen, keep up the good work!

Yawen, are there any interesting things you can tell us about your-- you're working on your dissertation and I know it involves creativity and AI.

Are there just some short findings or preliminary findings that you'd be willing to talk about?

Okay, thank you so much for the opportunity.

I have learned a lot from your perspective.

In my dissertation, I try to focus on the dynamic pathways of creativity

and the dynamic process of the collaboration between human-human collaboration and human-AI collaboration in the creativity test.

And I found so many differences, not only in the outcome, performance, but also in the process.

I found that human-human collaboration, they performance data,

they perform better in the convergent thinking and the human-AI collaboration

they perform better in the divergent thinking.

And in terms of the process, they have, there also different differences in the cognitive process and affective process.

So I think is an interesting area and I hope to do more research and have more advice from you all.

Do you have some patterns o research interest in it?

Well, that's a great place, I think, for us to conclude today, because we will be looking to you actually, in the future.

You're the next generation of scholars who are going to continue to bring us forward, Yawen.

And we are grateful for all the work you and your cohort of colleagues are doing.

Xiaoting, it is always a pleasure to learn about the practices, the possibilities, the research agenda of our colleagues in China.

We look forward to future cooperations with you, San Jose State and Peking University.

I know we have some ideas planned going forward.

I think we're going to do everything we can to keep each other just informed and

asking each other questions that we think are both philosophical and operational.

And with more guidance coming from other experts in the field, we'll have much more to say about this topic.

But today you got us started on thinking about process as the construct,

and really getting into the question of how assessment as a paradigm is being

shifted and changed and offering some very exciting new opportunities ahead.

We thank everybody for attending today. Sorry for any technical difficulties.

They came on my side here, unfortunately in Washington State.

But we'll do our best to catch up with the 21st century and the use of the internet here in the US.

Thank you so much, everybody, and have a great and wonderful day, and we'll catch up with you soon.