

Hello, and welcome to the final of five video chapters in us telling you about the I-SMART Goal Two Scenario-based Tasks project. I'm Bob Dolan, and I'll be discussing what we learned from the project.

So, the previous chapters described the process, and now we're going to be going over final conclusions. As a reminder, Goal Two of this I-SMART enhanced assessment grant was about designing, developing, and evaluating learning map model-based assessments. And this particular study within Goal Two focused on co-designing and evaluating testlets for what we call the secondary population of students. Those without significant cognitive disabilities.

We develop these scenario-based tasks to evaluate a range of depth of knowledge in science. We use deeper application of UDL principles, and we emphasize the formative use of these as instructionally embedded testlets.

So, I'm going to divide up what we learned into five sections. Going to start talking about the design process itself, which Kimberly described. We're going to talk about what we learned from the think-alouds and interviews in terms of student interaction and engagement with the tasks. We're going to talk a little bit about the design of the research study itself, and then we're going to talk about how these can be used to support formative assessment as a process, and also some considerations about what it would take to use these types of tasks in a balanced assessment system.

So, let's start with the design process itself. Some of our key takeaways is we had tremendous engagement with students in these co-design activities. Gave us an opportunity to learn what engages students, because they got to tell us what engages them in demonstrating their construct relevant knowledge, skills, and abilities, which by the way, did not include reading long passages. Students were very glad to tell us that during the design and actually during the think-alouds and interviews themselves.

So, they got to tell us because we asked them, "How would you like to show? If you were going to be evaluated on these skills and these knowledge elements, how would you like to demonstrate it?" And so inherently by giving them this opportunity, we learned what engaged them in the process.

The paper prototyping process itself was fantastic. We originally imagined that in the co-design process, we'd be developing these low and medium fidelity computer-based prototypes, but we decided to do it on paper. It allowed for much more rapid iteration, and then the process itself of working with the students was fantastic. I think mostly because we could work around a table with multiple students there on a flat surface, they could all participate. They could all watch.

It was a much more engaging task for them than if it was a student sitting at a computer, even with the other students sitting around them, even if we used a projector and a screen, it wouldn't have been nearly as interactive for a group of students.

Moving on to the actual student use of the testlets in the think-alouds and looking at their interaction and engagement. Looking at research question one, do students understand how to interact with the new item types? Yeah. Students were really clear on this, largely because we worked very closely with students in the design. Students are pretty savvy with these interfaces. It was not an issue to have them understand how to use all these various tools.

The challenge, and we'll talk about this and more in a moment, was having them understand why and when to use some of these tools. In terms of the explicit choices we gave students, such as in choosing which ecosystem to study and which sources of evidence to use in building their argument, students loved it. A couple of quotes here that we went over earlier. "I like projects where you can choose what you want to do and not be forced to do something that you wouldn't want to do." "People might know about this one ecosystem, but other people might not know or not be interested in the other one as much," probably referring to themselves.

So great. Students loved it; it was engaging. Would it actually make a difference in terms of their ability to demonstrate their construct relevant knowledge, skills, and abilities? We can't really tell from this. We know engagement obviously is a gating factor in accurate assessment, but the degree to which, we can't tell from this initial study.

Specifically, we're curious about whether this would just be a novelty effect also. Students love this idea being given a choice, but if this was a regular thing, would the novelty wear off? And also, for some students, this might be additional cognitive load. It definitely is another element to the task, and it's going to take time and energy to do this. For some students, it might not be an effective paradigm.

What about implicit choices? So, this is students' choice of using the text-to-speech, the speech-to-text, the hints, the glossary, the Pros & Cons Clipboard, which was always available. This was harder. We didn't do a great job here, but it's not surprising. We didn't give students the opportunity to practice with these things.

We noticed times when it probably would've been effective for the students to use some of these tools. It would've helped them, again, access and demonstrate their knowledge, skills, and abilities tied to the constructs, but they didn't choose them. So, it raises a big question about how do we address students' varying metacognitive and self-regulated learning skills?

I was reminded when I was young and playing pinball machines, one of my favorite things, these vintage pinball machines always had this little set of instructions in the lower right-hand corner under the glass, and I'd always try to read them. And I was a good reader and I love reading instruction manuals, but I could never make sense out of these instructions. I couldn't tie the words to the actual task.

So generally, I would play the game and sometimes I'd go back to the instructions, and sometimes instructions would make a little bit more sense. This is a reminder to me that for students, whether or not we have explicit instructions or give students explicit opportunities to practice using some of these tools, it's going to be hard for them to integrate the information about how to use these tools and why to use these tools with their needs in the moment.

UDL has a lot to say about this though. Expert learners have that metacognitive self-regulated learning skill set to make these kinds of decisions. Expert learners in UDL are described as purposeful and motivated, resourceful and knowledgeable, strategic and goal oriented. And specifically, that last category, strategic and goal-oriented, which pertains to students' executive function and self-regulated learning skills.

Obviously, this is where we need to be spending our time with students, giving them these skills in general, as a part of instruction so that they will be ready. When completing tasks like the ones we are proposing, they'll be able to use the tools effectively.

So, the bottom-line is there's no quick fix for this. I'll talk a little bit later; we didn't give students much of an opportunity either to practice with these tools. We had them at the beginning of the task, play with them quickly, but clearly it wasn't enough and clearly it couldn't have been enough.

And I'll be honest, I keep finding myself falling into this, "Build it, and they will come," mindset. And this is yet another reminder that it's not that simple. These tools can help, but only to the extent that students really have had a chance to use them, understand them for themselves.

So, some recommendations in terms of the design of these tasks. Obviously, training and opportunity to use the tools instructionally and in practice is essential for their effective use during testing. Some specific things we've discovered, the Pros & Cons Clipboard, as you saw, was always available on the lower left-hand corner, I think it was.

There were times when we said, "Hey, now might be a good time to go to the Pros & Cons Clipboard," but they could ignore that. We believe that if a screen had come up, boom, right in front of them, "Here's the Pros & Cons Clipboard," you'd still give them the option to skip it, but it was there in front of them, that would've been better. So serial presentation of some of these tools would support students who might not otherwise choose to use these or might not see the value in using some of these tools.

Another observation, another suggestion is finer-grain corrective feedback. When students were completing the food web and there'd be three drop targets, and if they get any one of them wrong, we'd say, "Hey, you need to try this again," we could have given them some more specific information such as which two are correct, and which one needs to be corrected. That would hold them back from getting into this perseveration loop of just guessing.

Another consideration, this is tricky, deeper embedded supports because of the way the scenario was set up, and the argument building was set up, there were often double and triple, triple negatives, which were conceptually difficult. So, there may have been waves of scaffolding students through this process. We think these supports would be construct irrelevant, but we're not a 100% sure here, but that would be an interesting area of further exploration.

And then finally, very important, the hints we provided, some of them were task clarifications. Some were actually construct relevant supports, big difference. We need to be explicit about this. So just saying hint doesn't provide a lot of information for students on why they might want to choose to use that hint.

What did we learn in terms of study design? Well, as I've said, students need more time to learn to support tools. We had a single session with these students. It was too much to pack in to this 45-minute or one hour session. Ideally, we would've been able to work with the students more over time, similar to how we worked with the students in the co-design process, to have them understand how these support tools might help them.

As far as the think-aloud protocol itself, that actually worked really well with one caveat. This was a very new set of designs, radically new for many students. We needed more time post hoc, where we used some of the time, very limited time, 10 to 15 minutes after the think-aloud to do a semi-structured interview. We had a little bit of time to say, "Hey, on this screen here, you did such and such. What were you thinking?"

We needed more time and specifically this would've been our opportunity to understand why didn't students use hints more? And for that matter, why didn't they use any of the other implicit support options nearly as frequently as we had hoped? The behavior and interaction pattern visualization is something we did kind of at the last minute during analysis, this ended up being really cool and useful for us as researchers trying to make sense of what happened.

It allowed us to compare the students and allowed us to get a lot of detail all in one place. So, we're going to be playing with designs like this more in the future, because we think it does a great bridging of making these very complex data-rich scenarios more intuitive for understanding.

So, what about supporting formative assessment as an instructional process with tasks like these? Key takeaway. Well, our research question three was, "Does the task assess a range of depth of knowledges?" Yeah, we designed it that way and we do believe we did successfully give students an opportunity to independently demonstrate their basic, and/or their higher order depth of knowledge.

So, if we look for example at this very simple depiction, as I said earlier, take this with a grain of salt. This is only six students. We have examples of students who did better with the higher order depth of knowledge aspects of the task than they did with the basic depth of knowledge, and vice versa. You could see that here as well.

We think most importantly, this could be really good information for the students and the teachers, depictions similar to this, perhaps greatly simplified, can help educators understand where their students are in terms of, again, the more basic depth of knowledge, the more advanced depth of knowledge. That needs to be a study in and of itself, how to take this very complex and rich information and make it usable, understandable and usable by students and teachers.

Finally, big question. Could a task like this... So again, we designed this primarily for use formatively. Could it be used to provide summative information, and as such, be a component in a balanced assessment system? Lot of questions we first have to answer here. I think the biggest one is this tension between task comparability and reducing construct irrelevant variance.

So, some of the tricks that we pulled out to support this as a good, informative assessment task is providing students with immediate feedback. Yeah, but what are the psychometric implications there? And again, this whole question of choice, whether implicit or explicit, what are the psychometric implications?

One could argue from 30,000 feet, that we had task comparability across the ecosystems. I'd be willing to try to defend that, but in terms of some of the other choice elements, certainly, the hints and the glossary, we'd have to be really careful here in understanding how these options impact construct relevance of the task, and don't for example, give away answers.

One possibility is to provide a training wheels off mode, to design a task scenario where when it's going to be used for more summative purposes, some of these supports and options are no longer available. And then of course, there's always the cost-benefit analysis. These types of tasks are very expensive to develop, evaluate, to administer, to score. Does it pay off in terms of improved validity and improved reliability, improved fairness? Unclear. That said, these tasks, if they have an instructional value then it may indeed be time and money well spent.

So final thoughts about this project. We looked at the importance of providing learner agency in the assessment process, right? So, by getting students involved in the actual design of these tasks, we learned a lot about what would engage them or engage other students in doing these tasks. I want us to push the envelope here. I want us to, as a fields, also consider what would it be like to give students increased agency in the testing process itself?

We're giving them some agency here in terms of some of the choices, but could we give them even more choice? Start them off by saying, "How can you demonstrate what you know on the topic of food webs? How would you like to do this?" So yes, I'm talking about even greater degrees of choice, to the extent that they improve learner agency and fairness. I consider this a baby step in that direction.

And then secondly, the big challenges here come from the fact that we're looking at reducing deep sources of construct irrelevant variance. The UDL framework is helpful in that sense, in demonstrating or identifying what some of these sources are. The fact that students get engaged differently, have different executive function. All these construct irrelevant factors that are deeper than just perception, language, and symbols, physical action, which is a domain which we normally work in, in the assessment world.

So, these are hard questions, but again, we're hoping that this study is going to provide folks with some initial baby steps in unpacking this problem, so that we can develop assessments that are more fair and accurate for a wider range of students.

With that, on behalf of Kim Ducharme and myself, we'd like to thank the entire research team involved in this component of the I-SMART project, folks from Cast and Atlas. We thank you for listening and hope you've enjoyed the series. Bye, bye.