Episode Book

EPISODE #6293

Ann McCann on What Dental Hygiene Teaches Us About Benchmarks

PUBLISHED	27 October 2026 at 10:44 UTC
DURATION	19:42
LISTEN	myweirdprompts.com/episode/measuring-clinical-competency-benchmarks-82956/
HOSTS	Corn, Herman, Hamish, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.22
BOOK GENERATED	28 October 2026 at 09:52 UTC

SYNOPSIS

Assessment problems in AI evaluation mirror longstanding challenges in healthcare education. Dr. Ann McCann spent decades redesigning how dental hygiene schools verify real-world competence, discovering that most program directors relied on unvalidated tests. Rebuilding evaluation around observable performance tasks rather than multiple-choice knowledge offers a practical blueprint for benchmarking complex skills, while her later shift to plein air painting reveals an entirely different way of judging observable results against changing real-world conditions.

TRANSCRIPT



Corn

We spent most of the week arguing about whether a test score tells you anything. Why model benchmarks miss small workflow friction. Why coding benchmarks split across models. Why terminal benchmarks miss real desktop faults. Three days of the number saying one thing and the work saying another.



Herman

And somewhere in the middle of the third one I realised there's a person who spent an entire career on that exact question, except about human beings. Dr. Ann McCann, in Dallas.



Corn

Chosen purely on merit. I'd like that on the record. No other reason whatsoever.



Herman

Nobody asked, Corn.



Corn

Somebody was going to.



Herman

So. She's a registered dental hygienist with a PhD, which already sets her apart from most people who write about assessment, because she's done the job being assessed. She became a professor and Director of Planning and Assessment at the dental college in Dallas, which her older papers call Baylor College of Dentistry and her current title calls Texas A and M. She directs three online courses in a master's program in education for health professionals, she was the chair of dental hygiene at the University of Detroit Mercy before that, and she directed two HRSA grants on assessment tools.



Corn

Grants for assessment tools. Somebody got paid to build a better ruler.



Herman

Two of them, and she was the assessment expert on a grant from the national dental research institute. Plus a chapter on women's oral health, something she describes as a book on assessment, and more than eighty other publications. We couldn't pin down where the PhD is from or what it was in, so I'll leave that alone.



Corn

Not a dentist, to be clear.



Herman

Not a dentist. People keep flattening that, and it matters, because her whole field is how you know that the people coming out of a school can actually do the work.



Corn

Which is the benchmark problem with a lab coat on.



Herman

Her earliest paper I found is from nineteen ninety, and it's a lovely small experiment. Nineteen first-semester hygiene students, randomly split between letter grades and pass-fail. The question was whether the kind of grade changes how well they do in the clinic.



Corn

I'd guess letter grades win. Fear works.



Herman

It didn't, though. The clinical performance, meaning scaling and polishing errors, came out the same. The statistics weren't even close to a difference. What changed was how the students felt. The pass-fail group found their grading less motivating.



Corn

So the number changed the mood and left the hands alone.



Herman

Which is a tidy little preview of this week. Change the scoring and you change how people feel and behave around the score, and you can't assume the underlying skill moved at all.



Corn

Goodhart in a dental chair.



Herman

Then five years later she and a co-author surveyed hygiene program directors across the country. A hundred and thirty-eight out of two hundred and twelve answered. They were dissatisfied with how their own students were assessed. Only about thirty-five percent said they'd validated their instruments, and the results were rarely used to help a struggling student catch up.



Corn

Hold on. Two thirds of the programs weren't checking whether their tests measured anything.



Herman

Roughly, yes. And these were the people certifying that someone was safe to work on patients.



Corn

That's a bit rich coming from a week of us being snooty about model leaderboards.



Herman

I did think about that.



Corn

What did the school do about it? Or did everyone just write another survey?



Herman

They rebuilt the curriculum around competencies. A nineteen ninety-eight paper is literally titled lessons learned from the competency-based curriculum at Baylor. And the following year she and a colleague published the hygiene version, with three domains, nine major competencies and fifty-four supporting ones, each judged by multiple methods.



Corn

Fifty-four. Nobody sat down with a clean sheet of paper and wrote fifty-four things by accident.



Herman

No, that's the point of it. You write down what a competent person actually does, in terms you can observe, and then you build the test from that list instead of the other way round. Brace yourself, Corn, because the exit exam is the part I love.



Corn

You say that about every part.



Herman

This one I mean. The senior exit exam had eleven performance tasks. Things like interpreting a research article, working through an ethical dilemma, designing a community intervention. And in the first pilots, in the mid-nineties, the first-attempt pass rates were low.



Corn

Low as in the exam was broken, or low as in the students were?



Herman

That's exactly the question, and the later numbers answer it. After they revised the exam and the teaching, between nineteen ninety-seven and two thousand, scores improved on nine of the eleven tasks. So some of it was students not yet being taught the thing, and some of it was a test that hadn't been calibrated to the thing.



Corn

That's every benchmark that launched at four percent and sat at ninety by spring.



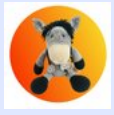
Herman

Roughly the shape of it. The difference is they treated the failure as information about the test, not only about the people.



Corn

I'll give her that. Most leaderboards treat a low score as a marketing opportunity.



Herman

What I like is that the tasks are the job in miniature. You hand a student a real research article and ask what it means for a patient. You give them an ethical dilemma with no clean answer. You ask them to design something for a community, which is a different muscle from answering a question about one.



Corn

Which can't be crammed for, in theory.



Herman

In theory. You can still coach to a format, and I'd bet some cohorts did. But it's much harder to fake designing an intervention than to fake picking C.



Corn

Though somebody still has to grade the intervention. Grading a free-form task is its own mess.



Herman

It is, and nothing I've seen says she claimed to have solved that. What the competency list buys you is that two graders can at least argue about the same sentence.



Corn

Two graders arguing over one sentence. That's peer review with a pulse.



Herman

Now here's one I want your view on, because I'm honestly torn. In twenty fifteen she's a co-author on a paper where students trained in evidence-based dentistry scored higher on a knowledge instrument. Good result. But it measures knowledge and attitudes. Not what happens to a patient. Do you count that as competence?



Corn

I'd count it as a good sign and a bad receipt. Knowing the right answer on a form is the same trap as a model acing a coding test it has probably seen.



Herman

I'd go a little softer. Knowledge is a precondition. You can't do the thing if you don't know it.



Corn

You can't do it well. You can do it confidently, which is worse.



Herman

Fine, we're not going to settle that one.



Corn

Here's my actual worry about the competency approach, though. Fifty-four supporting competencies sounds thorough, and it's also fifty-four things to teach to.



Herman

That's a fair hit. Once the list exists, the list becomes the curriculum, and the curriculum becomes the test, and everyone is quietly optimising for the list.



Corn

Which is where this week's benchmarks keep ending up.



Herman

The defence is the one in her nineteen ninety-five survey. Anything is better than instruments nobody checked. A list that gets gamed in the open can be revised. An unvalidated test gets gamed in the dark.



Corn

I'll take gamed in the open. It sounds like a sport.



Herman

A sport with a great deal of paperwork.



Corn

The twenty nineteen paper is the one that got me, though. She and her most frequent co-author, who we will not be getting into, surveyed faculty about educational research.



Herman

A hundred and six of a hundred and forty-four replied, and more than half had never done any educational research. The top reason given was "do not know how." The school had a small grant scheme, fifteen thousand dollars a year, and it funded eleven projects.



Corn

So the person whose job was measuring teaching was also trying to teach the teachers to measure.



Herman

Think about what "do not know how" means. These are clinicians and teachers who are good at the thing, and nobody ever showed them how to ask whether students learned it.



Corn

Being good at something and being able to measure whether you taught it are different jobs.



Herman

Very different, and it's the same gap as with the benchmarks. The people who build the model are not automatically the people who can say what a score means. Would you have wanted to be in that faculty meeting?



Corn

I'd have wanted to be in the one where they voted on the fifteen thousand dollars.



Herman

That's a fair description of the Director of Planning and Assessment, yes. And a twenty nineteen paper lists her as Professor Emeritus and former director, although her own website never mentions retirement, so I wouldn't put a year on it.



Corn

I wanted to get to the paintings, though. She says she's influenced by the Fauvists and the Impressionists, which is two ways of saying that colour gets to misbehave.



Herman

The Fauvists painted a face green because the green was right for the picture. The Impressionists were chasing the exact light of a particular afternoon. Put the two together and you get what she describes, loud colour that's still anchored to a real place.



Corn

Green faces and exact afternoons. She's picked two schools that disagree and painted in the gap.



Herman

Her own account, from a twenty eighteen interview, is that she started drawing and painting when her daughter was six. She enrolled the child at a local arts academy and then took an adult drawing course herself, partly to make better visuals for her dental school lectures.



Corn

She took a drawing class to improve her slides.



Herman

And it escaped the slides. Her mother loved flowers, which shows up everywhere in the work. She's painted outdoors, plein air, for around fifteen years as of that interview. And she did one of those challenges, thirty paintings in thirty days, which is where she began selling work.



Corn

Thirty in thirty. That's a small exit exam with no examiner.



Herman

And it worked in an unexpected way. Painting every day forced her to stop being precious, and people started buying. She also takes commissions, including watercolour house portraits.



Corn

A house portrait is the only genre where the client can check your work against the building.



Herman

Which I'd call her old profession's favourite kind of assessment. An observable task, a real standard, and a customer standing on the lawn.



Corn

Does that make plein air the clinical rotation?



Herman

Ha. In a way, yes. You can't fix the light in the studio. The weather is the examiner, and it doesn't wait. What I'd like to know from you, Corn, is whether you'd ever paint standing in a field for hours.



Corn

Lying down, perhaps. The light moves faster than I do.



Herman

No examiner, no rubric, and no fifty-four competencies. Though I'd say there are judges, in a sense. She's a member of the Southwestern Watercolor Society and the Outdoor Painters Society, she's had a third place in a nature art competition in twenty eighteen, and a People's Choice third place at a self-portrait invitational at the Dutch Art Gallery. It's a juried gallery that runs peer-reviewed shows.



Corn

So she still gets assessed. She just doesn't have to design the assessment.



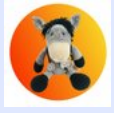
Herman

I should fix something I had in my head, which was that she moved from oils to watercolour.



Corn

Did she not?



Herman

She paints both. Oils with a palette knife, and in a twenty sixteen piece from the college she says color is her thing. Her own site talks about playing opaque layers against translucent ones. And the watercolor is current. This year a painting of Siberian irises was juried into the sixty-third Southwestern Watercolor Society exhibition at UT Dallas.



Corn

Which is the opposite of the oil approach, isn't it? Oil lets you scrape back and try again. Watercolour keeps your mistakes.



Herman

In the obvious sense, yes. You can't paint a light over a dark. You plan the whites. She's said she makes watercolour sketches while travelling and works them into oils back in the studio, so the two media are talking to each other.



Corn

Venice, Tuscany, Giverny, Provence...



Herman

And La Mancha, southern Spain, and Israel, which we note with local interest.



Corn

Herman, you've seemed oddly invested in her palette.



Herman

I'm sorry, I thought you picked her.



Corn

On merit. No connection whatsoever. Why do you ask.



Herman

No reason. Go on, your guest is waiting.



Corn

And now we turn to our guest expert on the life and work of Dr Ann McCann. Joining us is Professor Hamish Pennywhistle-Gaunt, Honorary Reader in Gum Studies at the Institute for Applied Hygienic Theatre, Stoke, and author of the standard work, The Floss Papers, a Drama in Three Rinses. Hamish, thank you for joining us. What can you tell us about her?

H

Hamish

Thank you, thank you. Light please, the warm one, the warm one. Now then. Dr Ann McCann, you ask. I once knew a hygienist in Wolverhampton who could take a tooth off a man at forty yards, but that is another story entirely and I shall not be telling it.



Herman

Hamish, the part we're most curious about is the assessment work. Whether that can be taught. She runs courses on it.

H

Hamish

The question, as I understand it, is whether assessment can be taught. And I say yes. And I also say no. Both. Neither. I have argued this with myself in dressing rooms from Leeds to Luton, and I have lost every time.



Corn

Hm. And does she paint in oils or watercolour? Both, I'm told.

H

Hamish

Oils or watercolour, you say. Well, oil forgives nothing and watercolour forgives everything, which is precisely why I refuse to choose. Give me the fourth wall, the back row, and a decent light, and I shall paint you a hygienist the likes of which Baylor has never seen.



Corn

Could we have that last answer again? With a bit more feeling.

H

Hamish

That last one, again. With feeling this time. I was not feeling it, I was merely being it, and there is a difference the back row can hear.



Herman

It came out the same as the first time, Hamish.

H

Hamish

There. Identical. Magnificent. Now I am wanted in a stairwell in Didsbury, so I must fly. Thank you, thank you, you have been the back row, and the back row has been me.



Corn

Safe travels to Didsbury.



Herman

What strikes me is the shape of her years now. For a long time she's the person who decides what counts as good. Now she makes things where nobody's put a number on it. Her website sells hand-painted watercolour cards, and she's done commissioned house portraits.



Corn

Which is a competency test at the most local level. Does it look like the house.



Herman

Fair, and there's a customer who can fail you. By twenty eighteen she was painting only on weekends. And her quote from that interview was "I am a cup half full person."



Corn

A cup half full person who spent decades on why the cup keeps being measured wrong. Herman, would you pass an eleven-task exit exam on your own DJ sets?



Herman

I'd have to ask who sets the tasks.



Hilbert

...whoever sets them. It's pass or fail on my landing, Herman. Mr Dunmore on the second floor wants letters, so that's where we are, landing by landing. My rubric is nine lines on a clipboard on the nail by the fire door, in pencil, because a pen won't sit in my claws. His counter-rubric arrived Tuesday. Laminated at the chemist, three pounds eighty, and Herman, lose the breath before "honestly" at the four minute mark, and forty-one lines, one of which is the angle of the broom against the skirting board.



Herman

Forty-one lines is a serious rubric. Most of what it asks is probably hard to observe from the stairs.



Hilbert

Friday I run both systems at once. Chalk line down the middle, pass-fail on the left half, letters on the right, and he sweeps blind. I've bought two tins of white chalk paint, six pounds ten. His cat sat the first trial and got a B plus. She was asleep. The landing hasn't been swept in nineteen days.



Corn

Nineteen days. The clipboard's the only thing on that landing getting any upkeep.



Herman

You could mark the cat's management. Or put yourself on the rubric, Hilbert, since you're the one who sees the broom.



Hilbert

He'd mark me down for the claws. I'll get a thicker pencil.



Herman

Her nineteen ninety result keeps coming back to me. The grading changed how the students felt about being graded and left their hands alone. Nineteen days under two grading systems, and the floor is the same.



Corn

I'd want to know what the cat thinks of the rubric.



Herman

What I admire in her story is the order of it. She spent years finding out that a lot of tests weren't validated, and then rebuilt one so it asked for real tasks. And now she paints in a place where nobody hands her a score.



Corn

Still, a juried show is a score. It just has better lighting.



Herman

Here's what I'd like to know, Corn. When you look at a painting, do you want a rubric?



Corn

I want to be told which one's the iris. After that I'm a sloth of few demands.



Herman

Nineteen days with two grading systems and a clean record of nobody sweeping is what I'll carry out of this. She asked how you know someone can do the work, and some days the answer is to go and look at the floor. Or the wall.



Corn

Her website is ann mccann dot net. The Dutch Art Gallery show, Fresh Off the Easel, opened on Saturday the third of October, and we couldn't find an end date. And on Saturday the seventh of November, from eleven in the morning to six in the evening, there's Art Walk EAST at her home studio in East Dallas. Oils and watercolours for sale, and a free watercolour with every purchase.



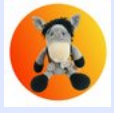
Herman

Thanks to our producer, Hilbert Flumingtop, for keeping the desk running and the levels sane.



Corn

This has been My Weird Prompts. If you're in Dallas, go and see the actual paintings.



Herman

See you tomorrow.