Hi. I am **Don Glass**, the Director of Evaluation and Field Work at the **National Commission for Teaching and America's Future** (NCTAF), an national education non-profit based in Washington, DC. This presentation is titled **STEM Curriculum Design and Evaluation Tools**

There is a growing consensus among school reformers that the implementation of the Next Generation Science Standards (NGSS) and the Common Core State Standards (CCSS) will require in-depth intellectual engagement and ongoing practical exploration of how cross-disciplinary content is translated into instruction for diverse learners. These new standards prompt for assessment and curriculum that make connections across STEM, ELA, Social Studies, and Arts practices and content. This requires educators from across subject areas to collaboratively design curriculum together.

Can evaluators play a developmental evaluative role in supporting this learning design and feedback? We think so. This presentation examines a set of curriculum design and evaluation tools that provide structure and formative feedback for this complex, inter-disciplinary curriculum planning process. The tools were piloted last year with 25 professional learning teams of middle and high school educators in several school districts in MD.

This session will focus mainly on the methodology and format of the data collection tools, but will provide some snapshots of the rich formative data for monitoring by program staff, documentation for teacher-generated case studies on Project-Based Learning, and baseline data for additional focused work on assessment design and scoring.

As an internal evaluator for a non-profit school reform support organization, I am interested in **using the tools and processes of evaluation to improve the practice of teachers**. With funding from the Carnegie Corporation of New York, I have worked to design and test a series of data collection and analysis tools and protocols that would be helpful and useful for teachers. Here are some of the features of this evaluation orientation:

˝ **Embedded** evaluation information gathering and reporting in program routines and expectations
˝ Provision of design and evaluation **capacity-building** for program participants at their **level-of-use**
˝ **Grounded** data collection and use in relevant curriculum design and timely assessment feedback to **reduce evaluation burden**
˝ Teacher input and **expert-user feedback** in design cycles for tools
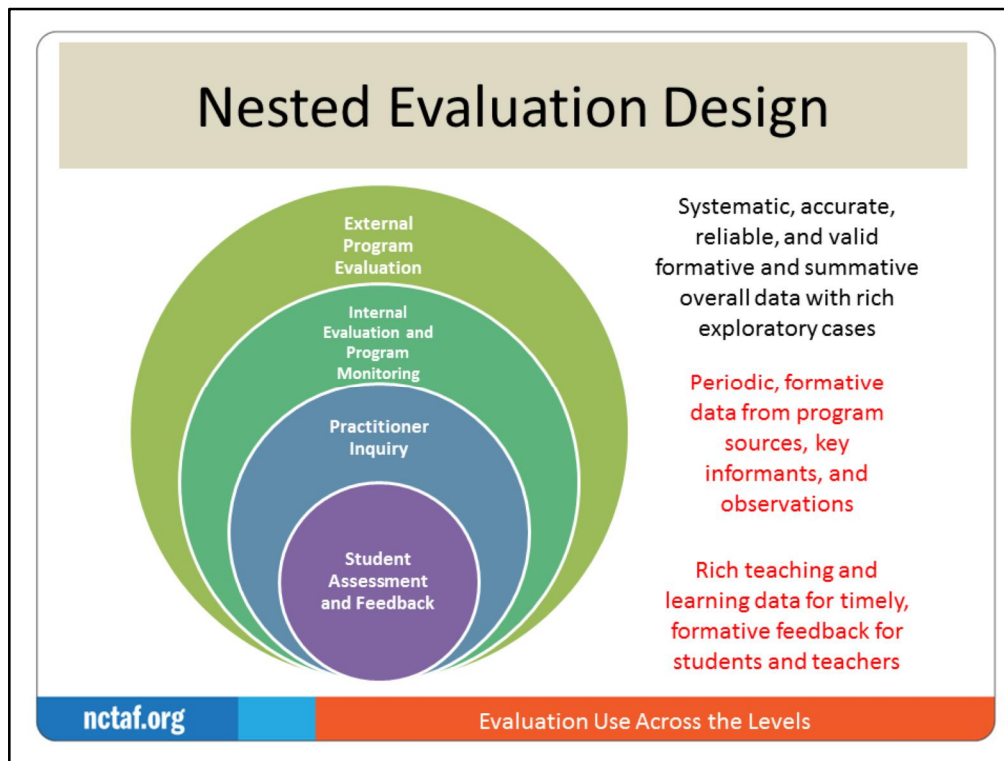
[Audio Describe Slide]

All of these nested levels can use evaluation concepts, methods, and tools to improve their work (i.e., students, teacher, program staff). **My role as an internal evaluator, required a shift to be an evaluation capacity-building coach** who works with practitioners during all phases of the project to use practical measurement to understand and improve practice.

You will see some **similarities to Participatory, Empowering, and Utilization-focused Approaches to Evaluation** (Fetterman, et. al., 2004, Patton, 2008). Similar to the steps Empowerment Evaluation, the tools and protocols helped educator teams to collaboratively take stock, set goals, and monitor progress for their collective work (project-based learning (PBL) units-of-study). The tools also helped building capacity for authentic assessment design. Several of the tools were embedded in reflective discussion protocols that prompted for the sharing and analysis of evidence of improvement, as well as for related further planning or action. The role of the evaluator was to validate the tools and data, build capacity for the use of the tools, and facilitate evidence-based feedback and action among teams.
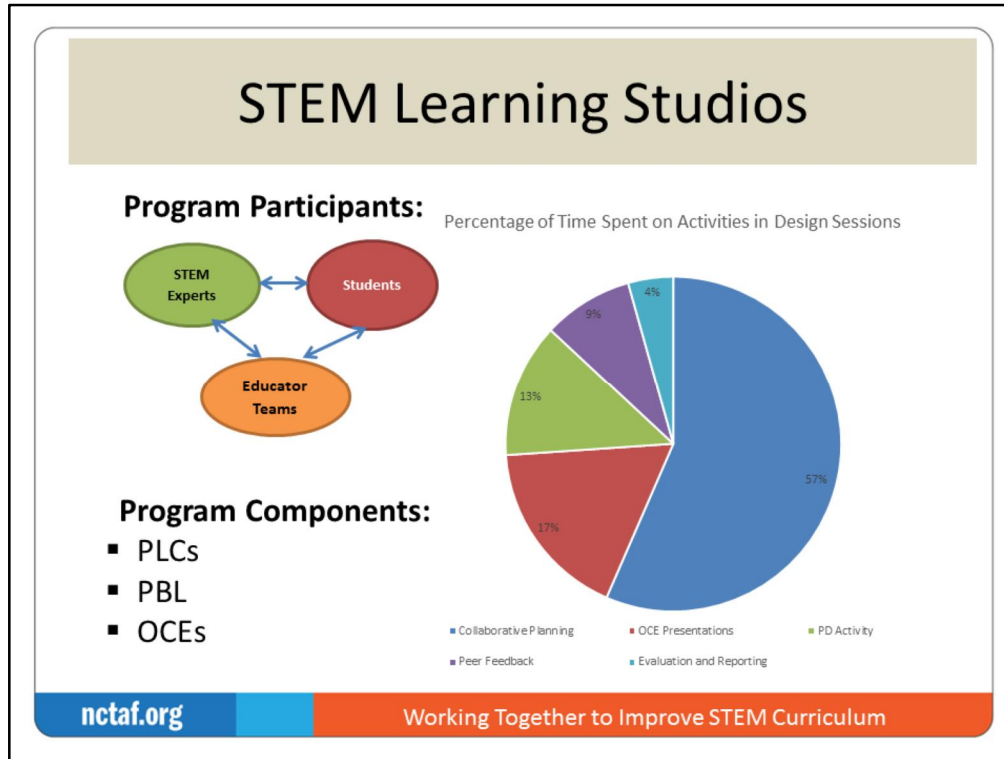
NCTAF STEM Learning Studios

The program provided professional development supports for cross-curricular professional learning communities (PLC) focused on the collaborative design of project-based (PBL) units of study with outside content experts (OCE) who provided content expertise, resources, and curricular coaching. Our *theory of action* is that a high-functioning PLC with an instructional design focus on PBL *and* content, resource and coaching supports from an outside content expertise can accelerate the rate of improvement of collaborative work and learning design.

NCTAF facilitated summer and quarterly full-day Design Sessions that provided collaborative planning time (57%), OCE presentations of content and resources (17%), curriculum design and evaluation capacity-building (13%), and peer and program feedback routines (13%).

STEM Learning Studios

Team Member	Evaluation Focus	Tools
Teams of EDUCATORS	• Collaborative Culture (PLC)	• Baseline and Interim Self-ratings • Collaboration Discussion Protocol
STEM Experts	• Curricular Improvement (PBL)	• Curriculum Map • PBL Checklist • Teacher Final Survey
STUDENTS	• Student Learning • Student Interest	• Student Retrospective Survey

Internal Formative Evaluation| External Third-Party Evaluation

nctaf.org   Internal and External Evaluation Working Together

This table shows how each participant group was supported through evaluation to improve their work. For the teacher teams and STEM experts, we focused on improving collaborative culture (PLC) and curriculum design (PBL). The tools labeled in red were designed, tested, and used to gather baseline and interim data to inform practice. The tools labeled in black were administered by our external evaluator WestEd as a summary evaluation on the focus areas--- which was especially helpful for getting summary student outcomes on learning and interest/engagement.
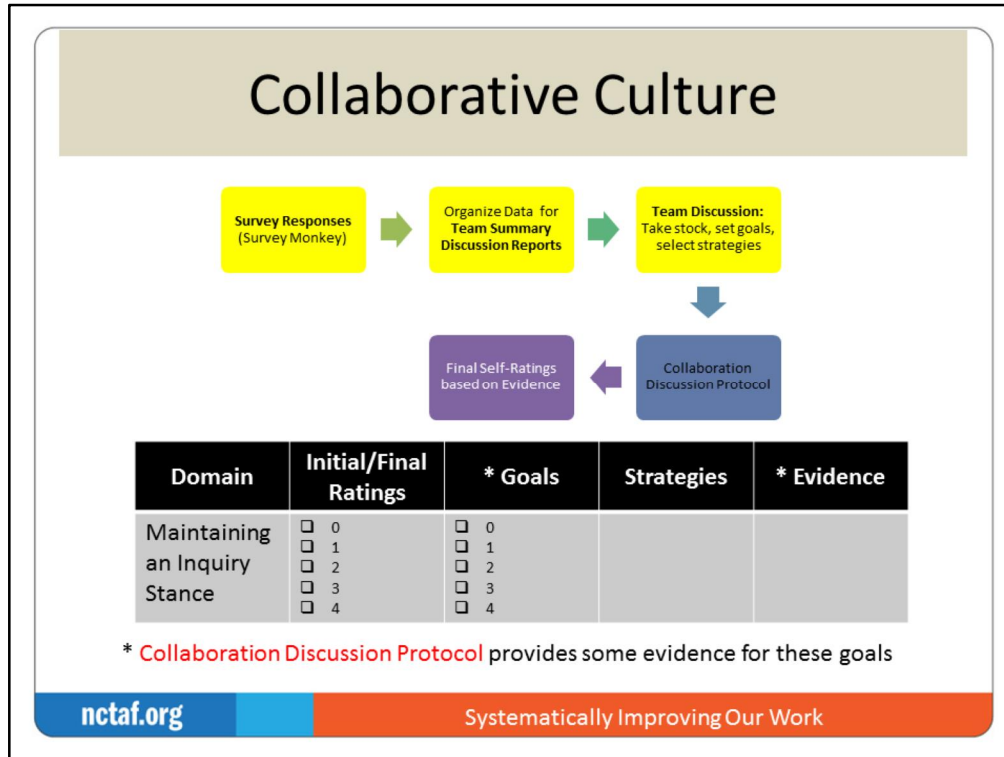
Collaborative Culture is one of the evaluation focus areas. The goal here was to improve the levels of collaboration of the cross-curricular teacher teams and OCE's, as well as increase their focus on improving curriculum and student learning. The basic flow of the evaluation work started with gathering baseline data in a survey with multiple items in six domains identified by several literature reviews. These scores were aggregated to the team level with some information of the range or spread of ratings on each item. A report of the scores were provided to the teams to discuss areas of strength, identify areas of growth, set some goals, and then propose some strategies to try out. Periodically at the quarterly design sessions, these goals and strategies were revisited. Evidence of growth was documented and discussed. At the end of the year the individual surveys were administered again and the results were presented to the teams to make new plans for improvement.

The survey instrument was adapted from a NCLE framework, survey tool, analysis report, and discussion protocol around Collaborative Culture that aligns well with the literature reviews conducted by NCTAF and its partners at Pearson.

This is a screenshot of some ratings items for the domain of maintaining an inquiry stance. Teachers completed this survey in Survey Monkey at the initial Summer Design session and the 4th quarter Design Session.

This is an example of the team survey results and a team's responses to the discussion protocol. The purpose was for teams to use data to self-evaluate themselves by taking stock, setting goals, exploring strategies, and then gathering evidence of improvement.

Collaborative Culture

Collaborative Culture Survey
Summer Design Sessions 2013

	# Teams	8	5	10	23
	Total Team Members	59	33	50	142
	SUM of Team Scores	17.06	16.08	18.97	
	PD Priority Ranking				
I. Deprivatizing Practice	1	2.60	2.41	2.89	7.90
We observe each other in the classroom and provide feedback to each other.		1.26	0.38	2.08	3.72
All members of the group stay engaged and accountable to each other.		2.56	2.52	2.65	7.72
We make commitments to try things in our classrooms and report back on the results.		2.90	2.60	3.13	8.62
We are comfortable sharing evidence about what is happening in our classrooms.		3.39	3.31	3.57	10.26
We share what we learn with others beyond our group.		2.56	2.73	2.88	8.17
Our group's work connects to the broader goals of the system in which we work.		2.94	2.91	3.03	8.88
IV. Using Evidence Effectively	2	2.63	2.43	3.04	8.10
Our collaboration stays grounded in evidence of student learning.		2.66	2.72	3.07	8.45
We have the skills in our group to use data effectively.		2.64	2.28	3.17	8.08
When we try something, we analyze the impact on student learning.		2.54	2.39	3.08	8.02
We examine and discuss student work with each other.		2.66	2.34	2.82	7.82

nctaf.org   Using Evaluation to Inform Our Work

This is a report that was run to get a summary of the average scores for teams across the districts. Conditional formatting was used to visually identify the strength or weakness of particular domains and items. Red is on the low end of the scale and green is on the high end of the scale. The sums of items were used to rank the priority of professional development or capacity-building supports. This information was used by staff to craft the capacity-building, tool design, and evaluation/feedback activities of the design sessions. In this case, peer observation and using student work evidence to evaluate work seemed to need the most support, which prompted the initial development in assessment design tools and looking at student work protocols.
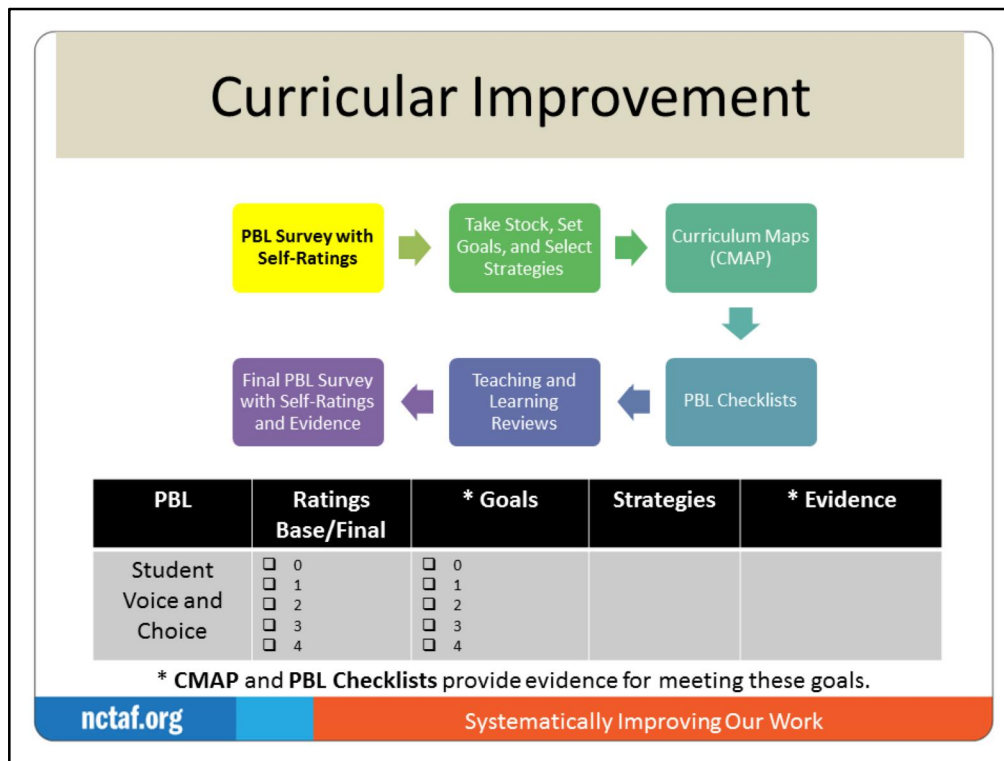
This is another example from the **curricular improvement evaluation focus**. In this case a **survey** was designed and administered based on the Buck Institute's PBL rubric to get baseline information about the team's PBL knowledge and use. Team level information was used to set some goals and identify strategies like increase student voice and choice, provide a more public audience, or make the assessment task more closely mirror a real world problem.

For curricular improvement we used additional data collection tools. We had teams collaboratively **map their curriculum**, and then periodically review their PBL units of study using a **PBL checklist** that was based on the PBL Rubric. In response to the Collaborative Culture survey results, we also introduced a **protocol** for reviewing teaching and learning. Like for the collaborative culture program area, we ended the year by individually taking the survey again to provide ratings and qualitative evidence of improvement. We then reviewed the results as teams.

Curricular Improvement

Curriculum Map

Understanding by Design (UbD)	Project-Based Learning (PBL)
DESIRED RESULTS	Significant Content
	Driving Question
	Need to Know
ASSESSMENT EVIDENCE	Revision and Reflection
	Public Audience
LEARNING PLAN	21c Competencies
	In-Depth Inquiry
	Voice and Choice

nctaf.org — Informing the Design of Tools with the Literature

So again, here are some of the conceptual frameworks for designing the tools (e.g., Understanding by Design (Wiggins and McTighe, 2005), and Buck Institutes' PBL resources).

This is a screen-shot of the survey items in Survey Monkey. This page gathered ratings for assessment evidence and asked a few key open-ended questions to get information on current knowledge and practice.

This is our Curriculum Map that features the NGSS dimensions for the desired results, the NGSS PE for the assessment evidence, and multiple columns for activities and resources across subject areas over time. It is basically an outcomes-based logic model or UbD template with the NGSS language embedded.

We then put this curriculum map into google docs to make it more **flexible and share-able**. Additional supports were added such as **hyperlinks** to NGSS resources/tools, and **just-in-time supports** like annotations to remind you what the aspects of an essential question are as you are working on it. Each subject area teacher had access to the shared planning document. The STEM expert and NCTAF coaching staff did as well. The Google doc revision feature allowed us to monitor changes over time. We also placed a **PBL checklist** in the same document to make it easier to complete the checklist and provide supportive evidence.

Curricular Improvement

		ALL	SCIENCE	MATH	TECH	ELA	SS	OTHER
	N=	87	37	17	12	9	5	7
Desired Results								
Significant Content	standards	3.62	3.48	3.82	3.58	3.67	4.00	3.86
	big ideas	3.55	3.42	3.76	3.33	3.78	4.00	3.43
Driving Question	relevance	3.14	3.19	3.00	3.00	3.11	3.40	3.29
	essential questions	3.06	2.81	3.06	3.17	3.33	3.40	3.29
	AVG	3.34	3.23	3.41	3.27	3.47	3.70	3.46
Assessment Evidence								
Revision and Reflection	feedback-routines	3.02	3.13	2.88	2.75	3.13	3.00	2.83
	reflection-learning-formative	2.96	2.93	2.76	3.00	3.25	3.00	2.83
	feedback-use	2.84	2.70	2.76	3.17	3.13	2.80	2.67
	reflection-project-formative	2.73	2.77	2.44	3.08	2.88	2.40	2.83
	reflection-learning-summative	2.63	2.57	2.44	3.25	2.63	2.20	2.67
	feedback-peer-critiques	2.58	2.43	2.53	2.83	2.78	2.60	2.33
	reflection-project-summative	2.54	2.33	2.38	3.27	2.63	2.40	2.50
Public Audience	evidence-based presentations	2.37	2.27	2.31	2.67	2.38	2.00	2.67
	authentic audience	2.16	2.03	1.94	2.50	2.25	1.60	2.67
	AVG	2.65	2.57	2.49	2.95	2.78	2.44	2.67
Learning Plan								
21st Century Competencies	collaboration-group	3.08	3.07	3.19	3.08	3.00	2.80	3.17
In-Depth Inquiry	rigorous	3.08	3.07	3.06	3.13	3.13	2.80	3.50
21st Century Competencies	problem-solving	3.06	2.93	3.00	3.25	3.13	3.00	3.17
	inquiry	3.01	3.10	2.50	2.91	3.25	3.00	3.33
In-Depth Inquiry	analysis	2.99	3.20	2.81	2.92	2.88	2.60	2.83
	inquiry	2.91	2.90	2.81	2.83	2.75	3.20	3.33
Voice and Choice	independent work	2.88	2.73	2.88	3.25	3.38	2.60	3.00
In-Depth Inquiry	evidence-based solutions	2.80	2.93	2.75	2.75	3.00	2.60	2.83
Voice and Choice	voice and choice	2.62	2.57	2.50	2.92	3.00	2.20	2.83
21st Century Competencies	collaboration-community	2.23	1.90	2.19	2.58	2.63	1.80	3.00
	AVG	2.86	2.84	2.77	2.93	3.01	2.66	3.10
	< MEDIAN (2.91)							

nctaf.org   Using Data to Inform Our Work

Again, an example of a visual report that helped NCTAF staff see where the hotspots were in terms of providing professional development support for PBL curriculum design. An area of focus for improvement, based on this data, is in assessment literacy, design, and use.

This new area is guiding our decision-making for our next steps. We are now drawing from **developmental evaluation** (Patton, 2011) and **improvement science** (Bryk, Gomez, and Grunow, 2011) to think about systematic development and use of NGSS assessments. How do these processes address a specific problem of practice? Where do the fit into a larger complex system? How can we support rapid and practical measurement of our change ideas with tests in multiple contexts? How can we share data that help explain variation and support reliability rather than just fidelity?

Hopefully I will be back here next year to present on this work.

These tools, some improvement case studies, and rubrics are in development as an online interactive resource that should be available in early 2015 for STEM/NGSS coordinators and coaches www.nctaf.org