Principled design and development for embedding assessment for learning in games and simulations, with illustrations

Steve Ferrara, Paul Nichols, Kristen DiCerbo, Emily Lai
October 17, 2014

Presentation in the 2014 MARCES Conference, *Technology Enhanced Innovative Assessment: Development, Modeling, and Scoring from an Interdisciplinary Perspective*

Research & Innovation NETWORK

NextGen Learning & Assessment

Overview

- What is assessment for learning?
- What are games and simulations?
- Background on embedding assessment for learning in games
- Two examples

Goals for the presentation and book chapter

- Present principled design and development principles for embedding assessment in learning games and simulations
  - In pursuing that goal we'll talk about design and development principles and adding in the fun
  - Variations of these principles have been used for SimCityEdu, Mars Gen One: ArguBot Academy, PATL, and CCSL
- Illustrate implementations of embedding assessment for learning into a learning game and an online learning, practice, and feedback system for Common Core speaking and listening

Authors

Research & Innovation NETWORK

What is assessment for learning?

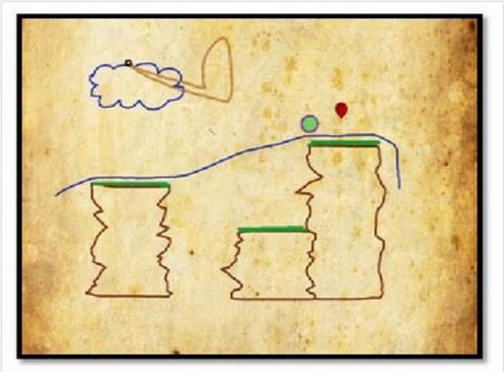
What is assessment *for* learning?

- “If formative assessment is about more frequent, assessment FOR learning…
  - Is about continuous…
  - informing the students themselves…
  - [telling]…what progress each student is making toward meeting each standard while the learning is happening—when there’s still time to be helpful”

(Stiggins, 2005, pp. 1-2)

AFL in gaming and simulation terms

- In games and simulations, we are able to gather evidence to inform our assessment models *while* students are engaged in a learning activity
- There is no need to stop the learning activity to take a test
- Goes by a number of names
  - Stealth assessment (Shute, 2011)
  - Ongoing, ubiquitous, unobtrusive assessment (DiCerbo & Behrens, 2014)
  - Invisible assessment



Newton's Playground

What are games and simulations?

Many definitions of games

- “A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome” (Salen & Zimmerman, 2003, p. 96)
- Serious games “have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement” (Abt, 1970)

Simulations

- Simulations can be thought of as "dynamic representations of spatial, temporal, and causal phenomena in … systems that learners can explore and manipulate" (Quellmalz, Timms, & Schneider, 2009)

- There are simulated *environments* and simulated *situations* that provide context and purpose for performance tasks

http://www.simscientists.org/sci_topics/index.php

https://www.ets.org/Media/Home/pdf/CBALMathSampleItems.pdf

Simulations (cont.)

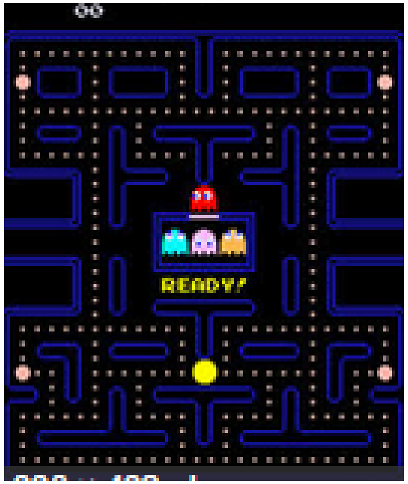
- Simulations provide reasons for users to explore functionality, manipulate phenomena, and pursue goals
  - Functionality, phenomena, and goals are intended to enhance authenticity and engagement and focus user attention in intended ways

Games and simulations: Contrasts

- Authenticity and engagement
  - Games lean more heavily on the engagement lever (e.g., narrative and rewards)
  - Simulations lean more heavily on the authenticity lever
  - Engagement elements can help particularly with significant retention issues (e.g., undergraduate engineering courses, struggling learners)
- Game mechanics
  - Actions that move the game forward (e.g., quests, resource management and, of course, points)
  - Game designers traditionally focus on how to use game elements to teach game play
  - Game mechanics can be tied to learning academic content
  - Simulations do not have this concept of mechanics to scaffold players' learning, although mechanics can be built in

Contrasts (cont.)

- Timing of challenges
  - Game designers are expert at offering the right challenge at the right time
  - Can be built into simulations as well
  - But concepts such as leveling up (e.g., getting from level 125 to 126 in Candy Crush…) are built into the game culture in a way that they are not built into simulations
  - Can capitalize on leveling up in simulations, as well



Background on embedding assessment for learning in games and simulations

Background

- Principled assessment design and development procedures
- Adding in the fun: Game design principles and procedures
- A principled approach to simultaneous assessment and game design: ECgD)
- Candidate topics for the book chapter that we won't address here
  - Designing-in assessment vs. ~retro-fitting
  - Creation and validation of measures
  - Deception
  - Automated scoring
  - Formative feedback
  - Reliability for games

Principled procedures: Common features

- Intended interpretation of results drives design and development decisions, from the first to the last
- Claims and evidence to support validity arguments are documented throughout, from design through operations
- Iterative design and feedback
- Thoughtful, not routinized design and development
- Use of content area expertise and research findings during design and development
- Use of reusable design tools (e.g., templates, models, design patterns) that promote efficiency and effectiveness gains
- Some differences
  - Use of templates to minimize variability of difficulty and complexity across tasks within templates vs. proactive manipulation of difficulty and complexity across tasks within templates

Principled procedures: Common features

- Intended interpretation of results drives design and development decisions, from the first to the last
- Claims and evidence to support validity arguments are documented throughout, from design through operations
- Iterative design and feedback
- Thoughtful, not routinized design and development
- Use of content area expertise and research findings during design and development
- Use of reusable design tools (e.g., templates, models, design patterns) that promote efficiency and effectiveness gains
- Some differences
  - E.g., use of templates to minimize variability of difficulty and complexity across tasks within templates vs. proactive manipulation of difficulty and complexity across tasks within templates

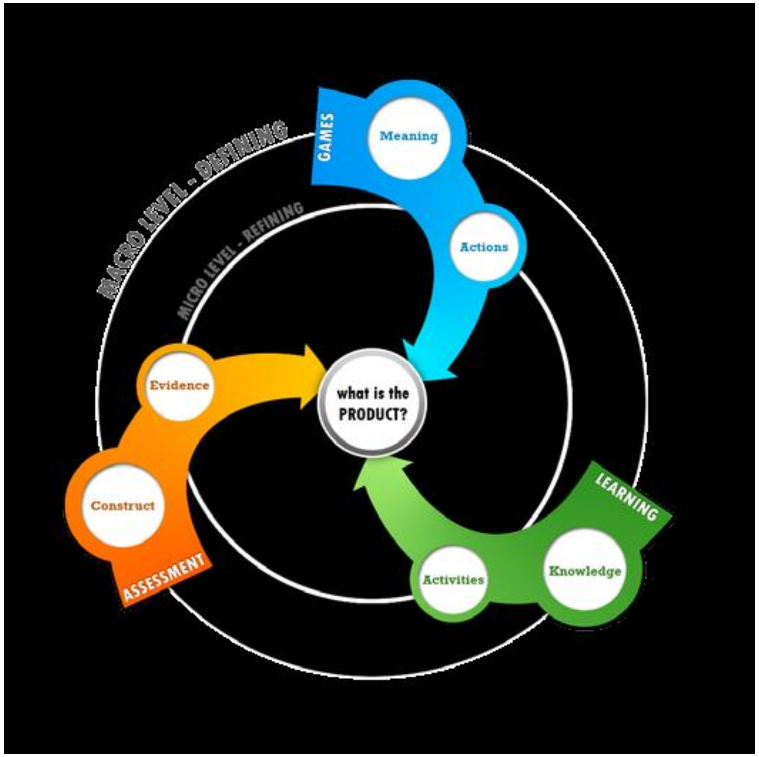
Adding in the fun: Game design principles and procedures

- As many theories of game design as there are game designers

- Key elements that need to be designed in:
  - Space – What does the environment look like?
  - Goals – How do you win? Can many people win? Are there multiple ways to win?
  - Characters and objects in the world
  - Mechanics – Actions that drive game play
  - Rules – What can and cannot be done?

A process for simultaneous game and assessment design

- Evidence-Centered game Design (ECgD) (Mislevy et al., 2014)
  - Optimized for embedding assessment in game design



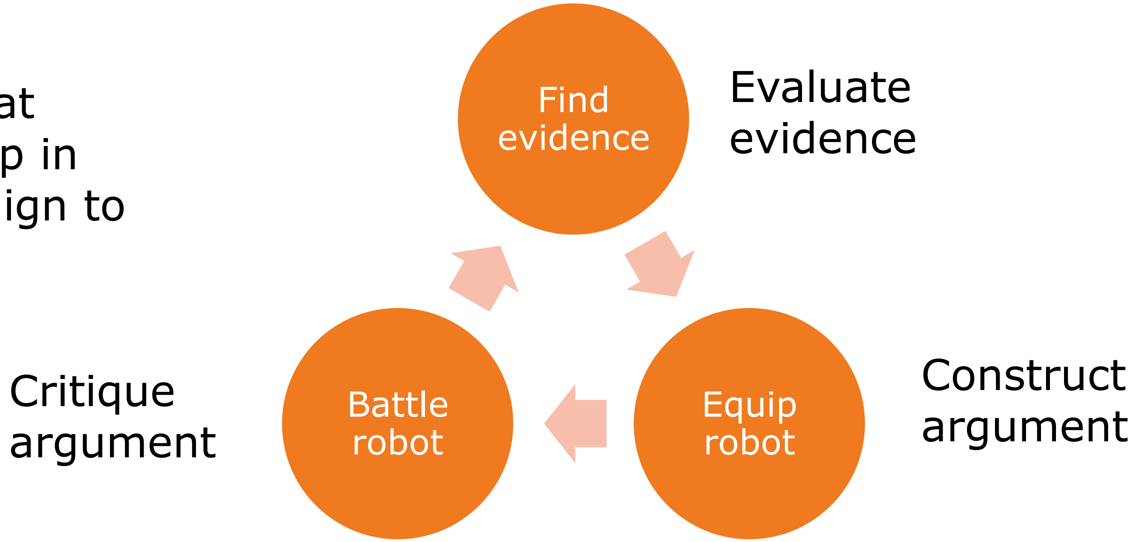
- Extension of ECD, which is optimized specifically for assessment design and development

- See Mislevy et al. (2014) for details on the process and tools

ECgD

- A process for creating video games that function as assessment and learning tools for academic competencies defined externally to the game
- Goal is to unify academically valued competencies with the goals of gameplay
- Bring game mechanics into congruence with learning goals

Example: Mechanics that form the game play loop in *Mars Generation One* align to the key elements of argumentation skill



ECgD (cont.)

- Two “state machines” (p. 126)—game state and assessment state—operate together to provide feedback to the gamer/learner
- Macro and micro design processes and documents, some which are familiar in assessment design; for example:
  - Domain modeling
  - Design patterns
  - “Augmented Q matrix”
  - Teachable agents
- Iteration: identify and make improvements based on gamer/learner responses
- Modularity and reusability

Other topics to address in the book chapter

- Designing-in assessment vs. ~retro-fitting
- Creation and validation of measures (i.e., evidence)
  - Game-based persistence (e.g., DiCerbo, 2014)
- Deception
  - Pitfalls in simulation (Behrens, DiCerbo, Ferrara, 2012)
- Automated scoring
- Formative feedback
  - Guidelines for generating (e.g., Shute, 2008)
- Reliability for games
  - Estimating score reliability for game contexts, data structure, and intended inferences (Nichols, Lai, Steedle, DiCerbo, & Ferrara, 2014)

Illustration: Research on Personalized Assessment, Teaching, and Learning (PATL)

Recap on PDE from the 2013 conference

Six Design Concepts

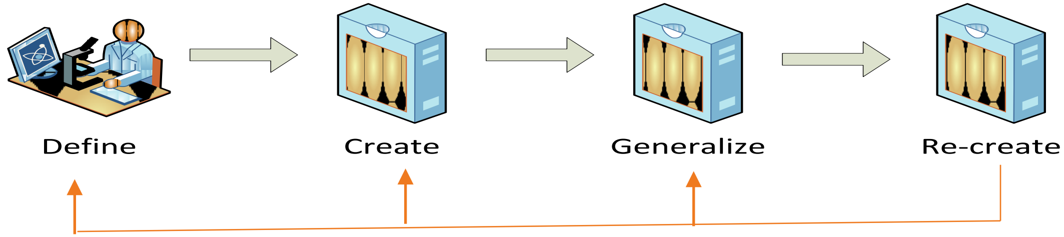
- **Construct**: What KSAs are you assessing?
- **Theory of Change**: What do you expect will happen when you assess? What are the mechanisms you believe will cause those changes?
- **Content**: How do you manipulate content to target those KSAs and mechanisms of change effectively?
- **Evidence**: How do you recognize use of the KSAs when you see it?
- **Communication**: How do you talk about what (and how) you will assess? (includes communicating with educators)
- **Implementation**: How do you work within practical constraints?

(Nichols, Ferrara, & Lai, 2014)

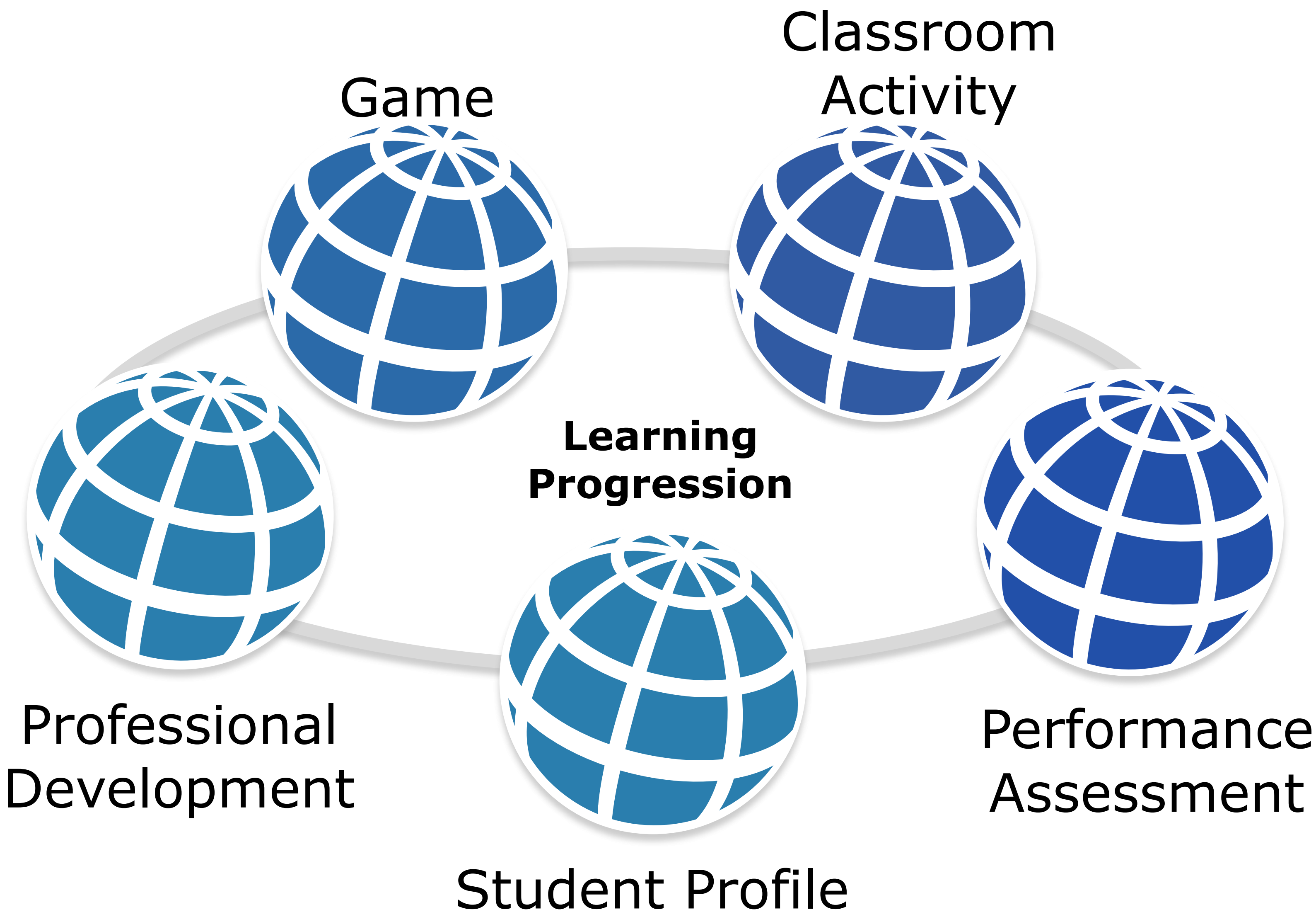
Brief overview of PDE

Four Stages of Design and Development

- Identify relevant learning sciences research to define and clarify KSAs, content features
- Create items and stimulus materials
- Construct reusable templates
- Use reusable templates to generate additional items and stimulus materials

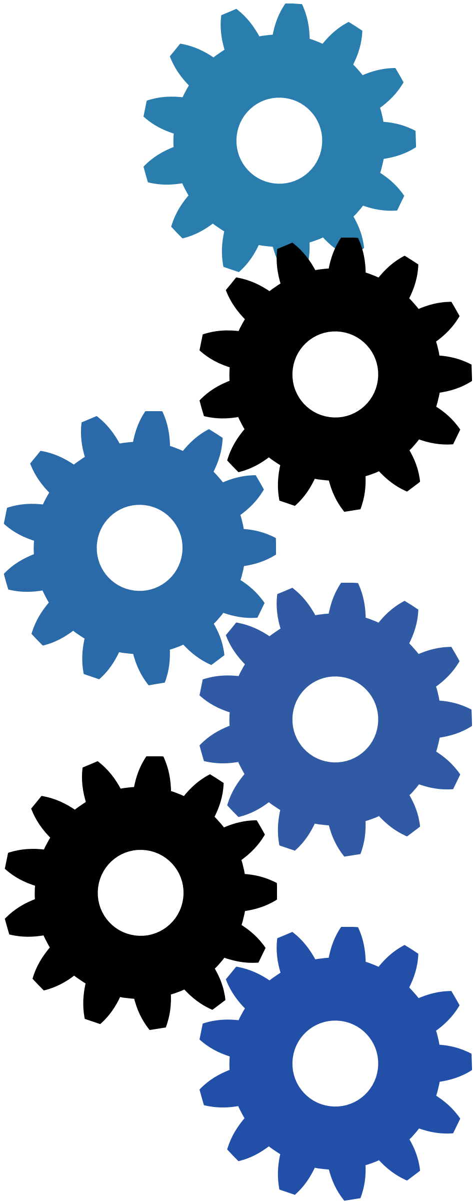


PATL Components



Description of the PATL Project

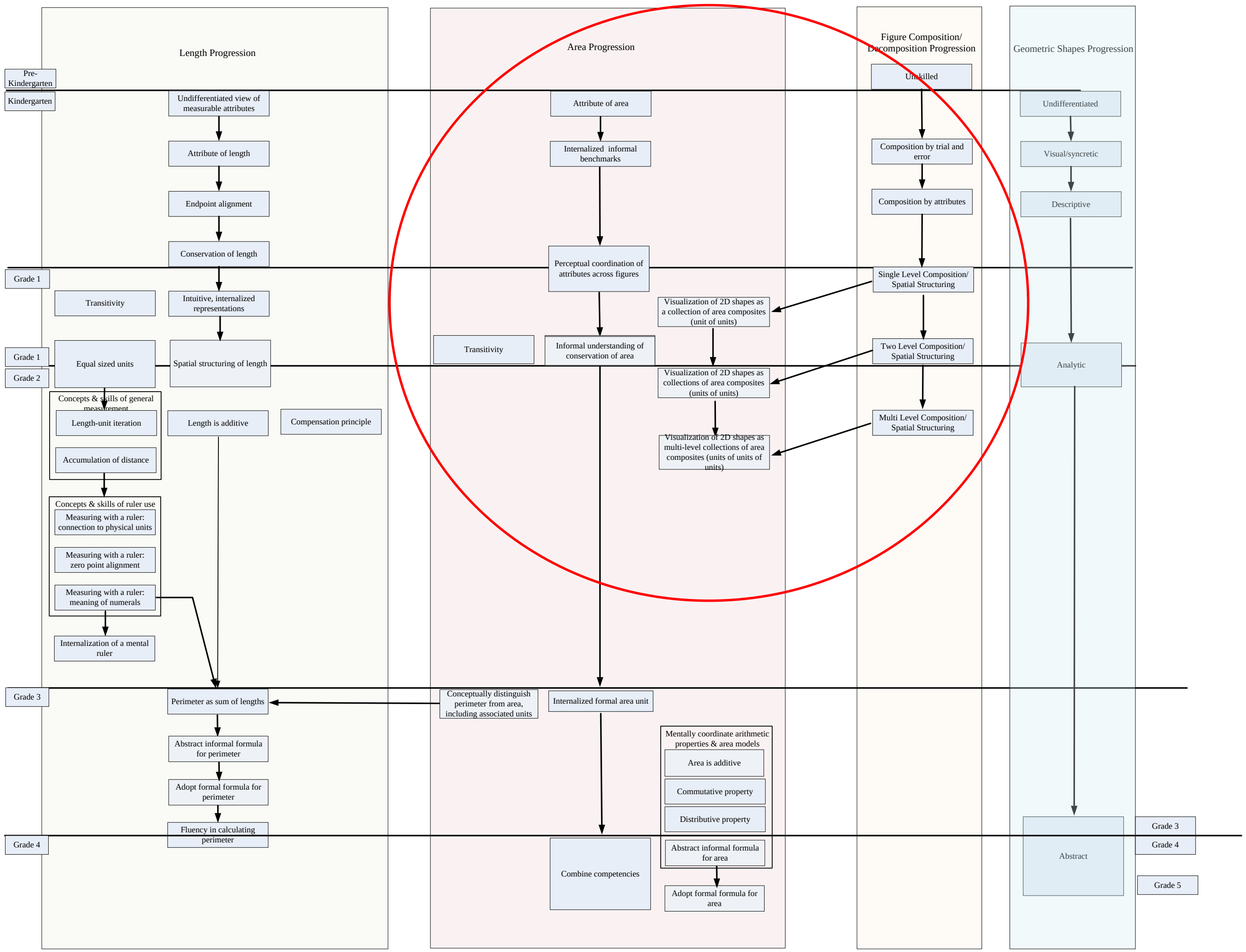
- Integrated learning system
- Teaching and assessing Geometric Measurement of Area
- Mechanism for integration is a shared learning progression
- Using Principled Design for Efficacy process to develop all components

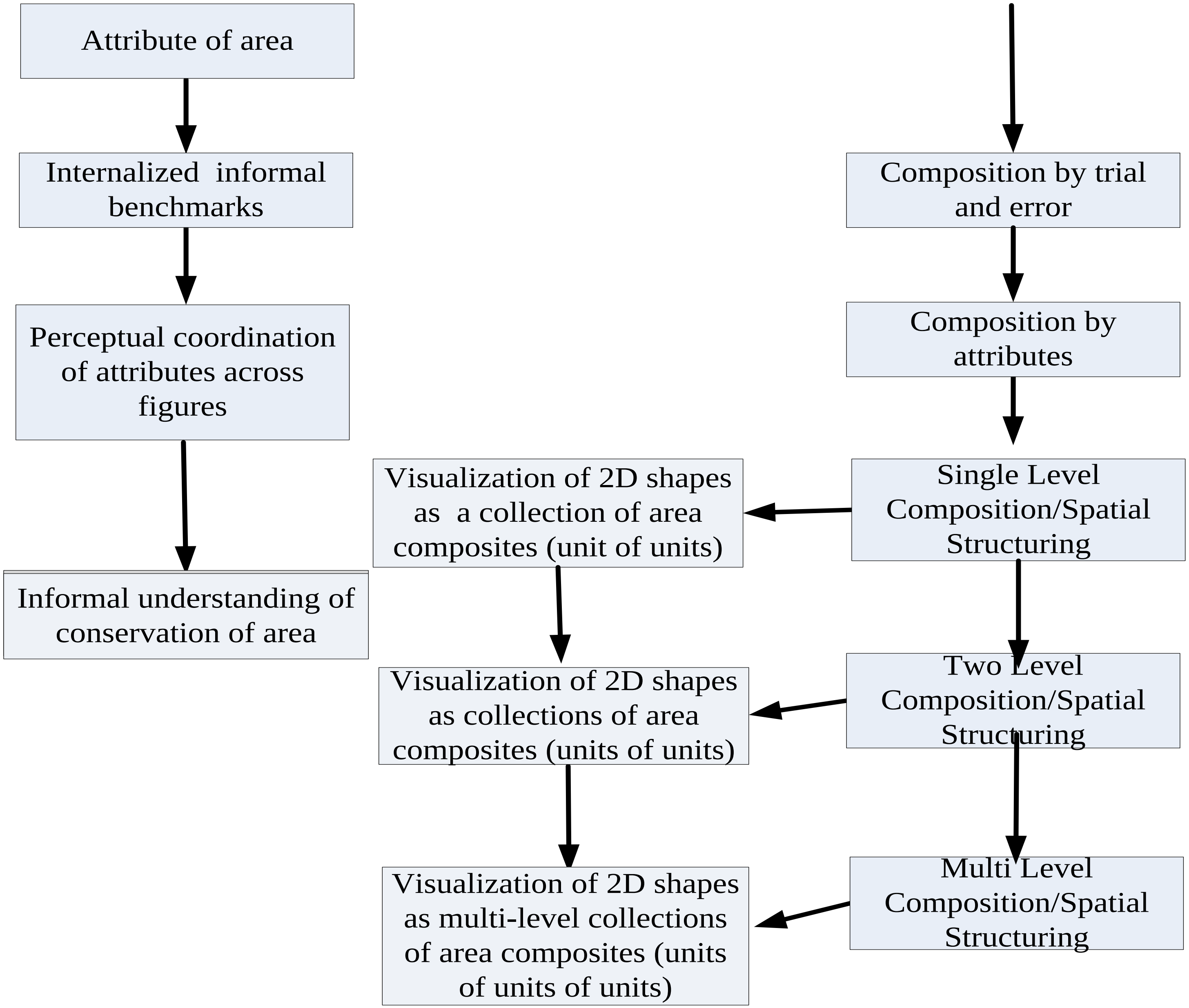


Stage 1: Clarify and extend the targets of inference

- Grades K-4
- Based on research from math education field
  - Progressions for the Common Core State Standards in Math (University of Arizona Institute for Mathematics and Education)
  - Doug Clements and Julie Sarama (SUNY, Buffalo)
  - Michael Battista (The Ohio State University)
  - Jere Confrey and colleagues (North Carolina State University)
- Used the research to extract levels that represent successively more sophisticated understandings and practices
- Mapped multiple strands (geometry, area, length)
- Identified
  - Performance indicators/behaviors
  - Common errors/misconceptions
  - Content features of tasks and stimulus materials

The area learning progression, grades K-4




Attribute of area
Internalized informal benchmarks
Perceptual coordination of attributes across figures
Informal understanding of conservation of area
Composition by trial and error
Composition by attributes
Single Level Composition/Spatial Structuring
Two Level Composition/Spatial Structuring
Multi Level Composition/Spatial Structuring
Visualization of 2D shapes as a collection of area composites (unit of units)
Visualization of 2D shapes as collections of area composites (units of units)
Visualization of 2D shapes as multi-level collections of area composites (units of units of units)

Stage 1: Identify and describe the features of content and performances

- Using the same research literature and the judgment of content experts
- Identify and describe content features likely to elicit use of learning progression understandings and practices at a stage
- Identify and describe features of performances that provide evidence of understandings and practices at a stage

Tool used to capture features of content and performance

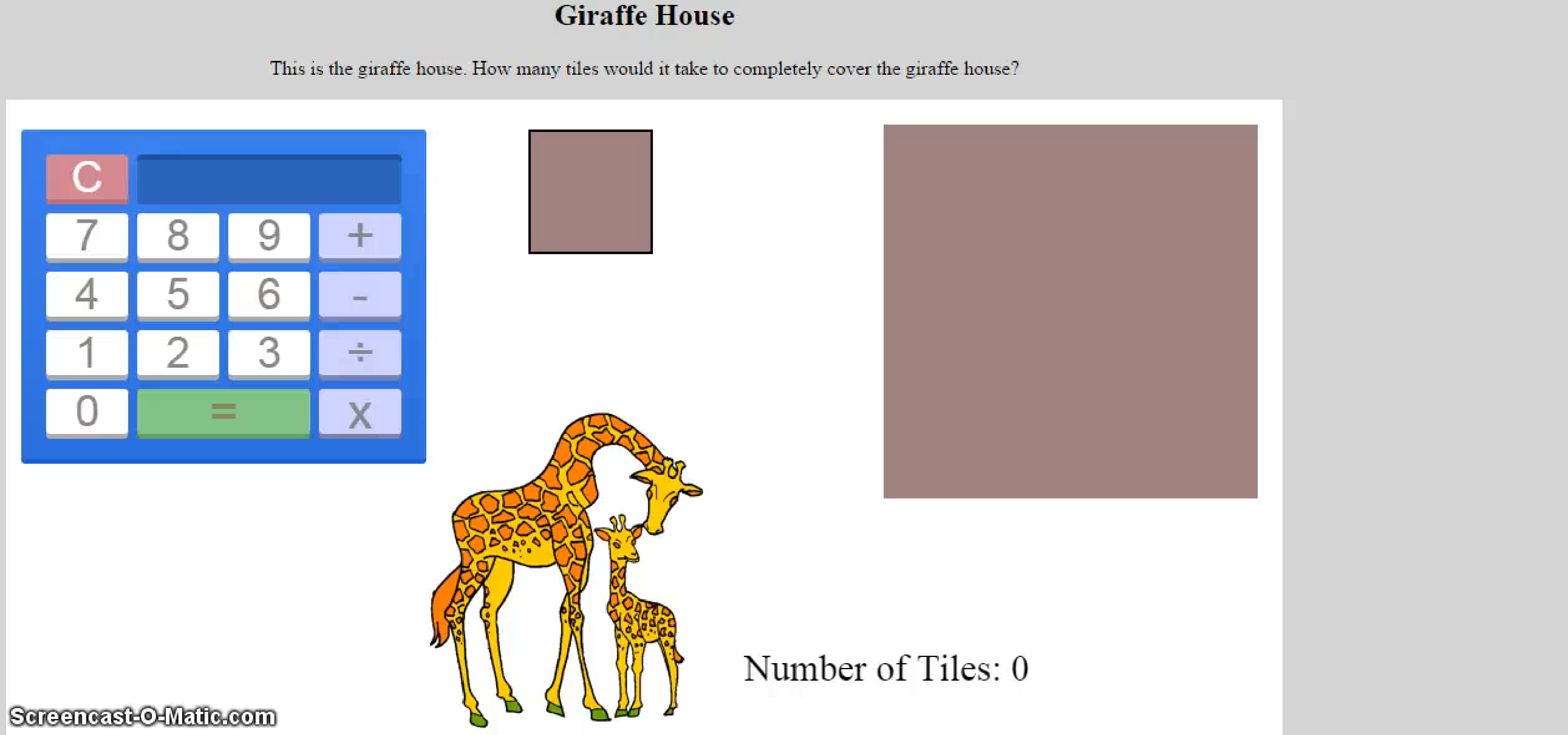
Feature	Source(s)	LP Stage	Evidence
Figures are drawn to scale	Joram, E. G., A. J., Bertheau, M., Gelman, R., & Subrahmanyam, K. (2005). Children's use of the reference point strategy for measurement estimation. Journal for Research in Mathematics Education, 36(1), 4 - 23. Barrett, J.E., Sarama, J., Clements, D.H., Cullen, C., McCool, J., Witkowski-Rumsey, C., & Klanderman, D. (2012). Evaluating and Improving a Learning Trajectory for Linear Measurement in Elementary Grades 2 and 3: A Longitudinal Study, Mathematical Thinking and Learning, 14:1, 28-54. Battista, M.T. (2004). Applying Cognition-Based Assessment to Elementary School Students' Development of Understanding of Area and Volume Measurement, Mathematical Thinking and Learning, 6:2, 185-204.	Area: Internalized, formal area unit Length: Internalization of a mental ruler	Without using measurement tools, student estimates with a reasonable degree of accuracy how many linear units or unit squares would cover a given shape, perhaps using fingers
Shapes are physically manipulable	Clements, D. (2011). Geometric and spatial thinking in early childhood education. In Clements & Sarama (Eds.) Engaging young children in mathematics. Mahwah, NJ: Lawrence Erlbaum Associates.	Area: Perceptual coordination of attributes Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition Length: Spatial structuring of length, length-unit iteration	Student places shapes side by side or on top of one another during direct comparison of area, student constructs or deconstructs composite figures of varying levels of complexity, student places units end to end along the length of an object
Level of scaffolding provided to help visualize internal structures of 2D shapes (e.g., presence/type of grid)	Battista, M.T. (2004). Applying Cognition-Based Assessment to Elementary School Students' Development of Understanding of Area and Volume Measurement, Mathematical Thinking and Learning, 6:2, 185-204.	Area: Visualization of 2D shapes as collections of area units, collections of area composites, or multilevel collections of area composites	Student counts each unit square to compute area, student skip counts the number of units in a row or column, student multiplies the number of units in a row by the number of units in a column
Flexible tool use (representing both more and less sophisticated strategies)	PATL project hypothesis	Area: Visualization of 2D shapes as collections of area units, collections of area composites, or multilevel collections of area composites, abstract informal formula for area Geometry: Single-level, two-level, and multi-level composition Length: Spatial structuring of length, length-unit iteration, concepts & skills of ruler use	Student uses units to spatially structure both 1D and 2D spaces, using unit iteration Students uses 1D and 2D structure to reason about area or length Students uses a ruler to measure length of objects and calculate perimeter and area
Shape attributes (e.g., regular vs. irregular, rectangular vs. non-rectangular, conventional vs. unconventional)	Clements, Wilson, & Sarama, 2004	Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition, visual/syncretic, descriptive, analytic, abstract Area: Visualization of 2D shapes as collections of area units, collections of area composites, or multilevel collections of area composites, abstract informal formula for area	Students recognize a wide variety of shapes by their attributes or properties and identify and describe those attributes Student combines shapes to make new shapes or pictures by trial and error, student uses shape attributes to combine shapes into pictures, student constructs or deconstructs composite figures of varying levels of complexity
Shape dimensions (keep side lengths within 1-10 to minimize routine counting, addition, or multiplication errors)	Battista, M.T. (2004). Applying Cognition-Based Assessment to Elementary School Students' Development of Understanding of Area and Volume Measurement, Mathematical Thinking and Learning, 6:2, 185-204.	Area: Abstract informal area formula, adopt formal area formula Length: Abstract informal perimeter formula, adopt formal perimeter formula, length is additive, fluency in calculating perimeter	Student counts each unit square to compute area, student skip counts the number of units in a row or column, student multiplies the number of units in a row by the number of units in a column
Shape orientation (conventional vs. unconventional orientations, e.g. presenting a square as a "diamond," triangles on their sides)	Clements, D. (2011). Geometric and spatial thinking in early childhood education. In Clements & Sarama (Eds.) Engaging young children in mathematics. Mahwah, NJ: Lawrence Erlbaum Associates.	Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition, visual/syncretic, descriptive, analytic, abstract	Students recognize a wide variety of shapes by their attributes or properties and identify and describe those attributes Student combines shapes to make new shapes or pictures by trial and error, student uses shape attributes to combine shapes into pictures, student constructs or deconstructs composite figures of varying levels of complexity
Shape dimensions given or not		Area: Adopt formal area formula Length: Adopt formal area for perimeter	Learner applies the formal area/perimeter formulas appropriately and with minimal prompts or cues. For example, when presented a rectangle with side lengths labeled but no internal structuring visible, the student retrieves and accurately applies the formulas.

Feature	Source(s)	LP Stage	Evidence
Figures are drawn to scale	Joram, E. G., A. J., Bertheau, M., Gelman, R., & Subrahmanyam, K. (2005). Children's use of the reference point strategy for measurement estimation. Journal for Research in Mathematics Education, 36(1), 4 - 23. Barrett, J.E., Sarama, J., Clements, D.H., Cullen, C., McCool, J., Witkowski-Rumsey, C., & Klanderman, D. (2012). Evaluating and Improving a Learning Trajectory for Linear Measurement in Elementary Grades 2 and 3: A Longitudinal Study, Mathematical Thinking and Learning, 14:1, 28-54. Battista, M.T. (2004). Applying Cognition-Based Assessment to Elementary School Students' Development of Understanding of Area and Volume Measurement, Mathematical Thinking and Learning, 6:2, 185-204.	Area: Internalized, formal area unit Length: Internalization of a mental ruler	Without using measurement tools, student estimates with a reasonable degree of accuracy how many linear units or unit squares would cover a given shape, perhaps using fingers
Shapes are physically manipulable	Clements, D. (2011). Geometric and spatial thinking in early childhood education. In Clements & Sarama (Eds.) Engaging young children in mathematics. Mahwah, NJ: Lawrence Erlbaum Associates.	Area: Perceptual coordination of attributes Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition Length: Spatial	Student places shapes side by side or on top of one another during direct comparison of area; student constructs or deconstructs composite figures of varying levels of complexity; student places units end to end along the length of an object

Stage 2: Use content and performance features to construct items and tasks

- Using the content and performance features tied to stages in the area learning progression, construct items and tasks
- For those items and tasks, create coding (not scoring) guides and reusable templates

Stage 2: Use content and performance features to construct items and tasks



Stage 3: Construct reusable tools and templates

- Have the content and performance features table
- Template that captures learning progression stage and content and performance features for the item
- Coding (not scoring) guide

Stage 3: Construct reusable tools and templates

Activity 10 Title: Giraffe House	
Brief Description of the Activity (this should indicate how this activity ties into the overall scenario or narrative of the PT)	This is the giraffe house. How many tiles would it take to completely cover the giraffe house?
Strand(s) of the Learning Progression that the Activity Targets	Area Figure composition and decomposition
Stage(s) of the Learning Progression that the Activity Targets	Figure composition and decomposition: Single-level composition Area: Using area units to measure area, Area unit iteration
Grade Level(s) that the Activity Targets	Grades 1-2
CCSS for Math	2.G.A.2 Partition a rectangle into rows and columns of same-size squares and count to find the total number of them.

Stage 3: Construct reusable tools and templates

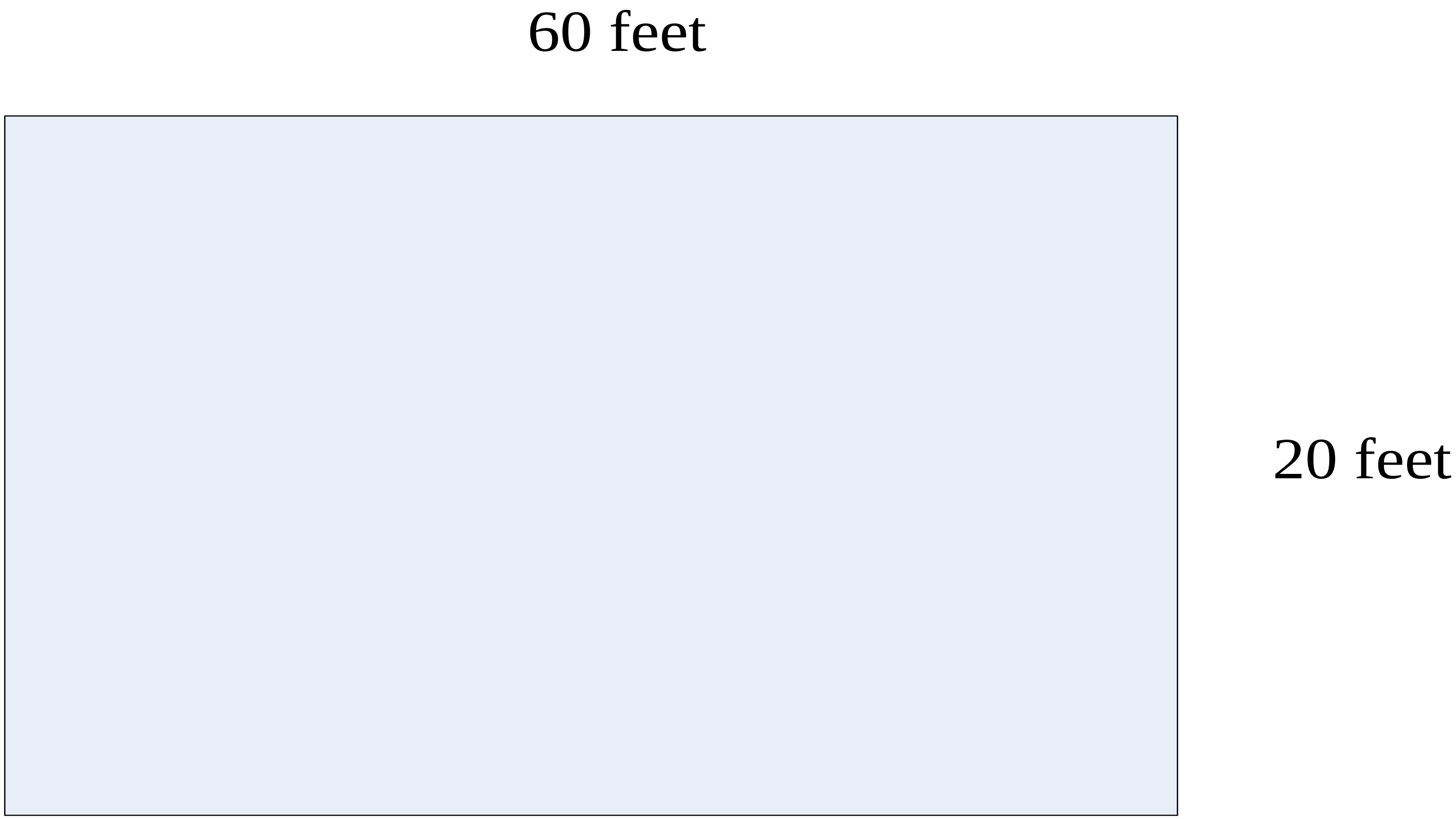
Evidence to collect (observable behaviors or potential student responses)	**Single-level composition: Student iterates individual unit squares to structure the shape** **Area unit iteration: In iterating individual unit squares, student does not leave gaps in between unit tiles** **Using area units to measure area: The student confirms the number of unit squares covering the bottom of the shape**
Content features	Shapes are physically manipulable, which allows students to compose or decompose composite figures and measure by iterating area units Shape dimensions: Side lengths within 1-10 to minimize routine counting, addition, or multiplication errors Shape attributes: Regular rectangle so area formula can be applied Level of scaffolding provided to help visualize internal structures of 2D shapes: Use of unit tiles allows students to visualize internal structure to support area measurement

Stage 3: Construct reusable tools and templates

Giraffe House Activity	Potential Student Response	Stage of the Progression
Part 1: How many tiles are needed to completely cover the giraffe house?	Student drags and drops tiles into the interior of the rectangle, but may overlap tiles or leave gaps in between	Single-level composition
	Student iterates area units, dropping tiles end to end and leaving no gaps	Area unit iteration
	Student submits the count of tiles as the total number of tiles needed to cover the shape	Visualization of 2D shapes as collections of area units (unit of units)

Conventional area task

A rectangular swimming pool is shown below.
What is the area of the swimming pool?

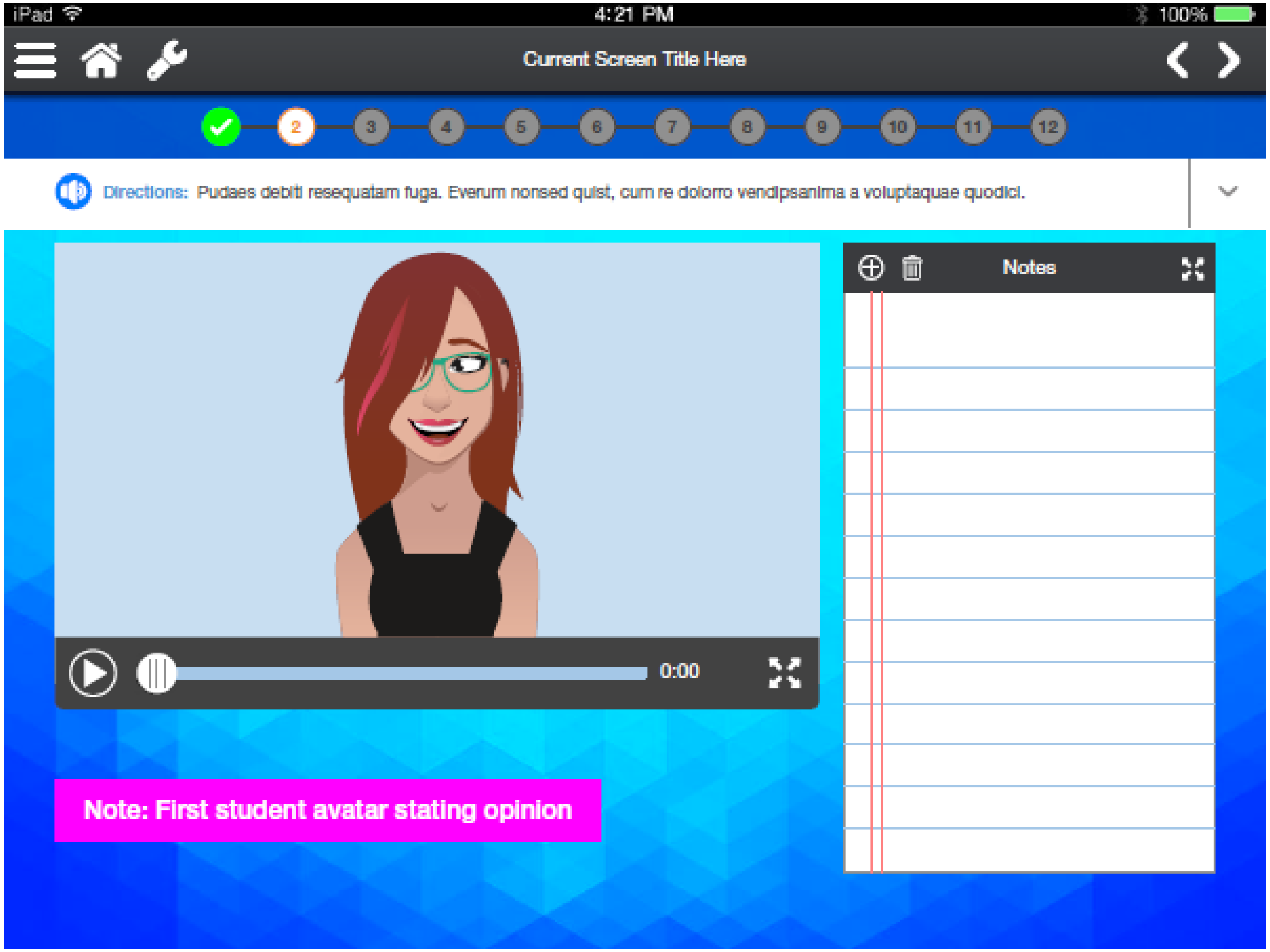


Thoughts about assessing learning and cognition

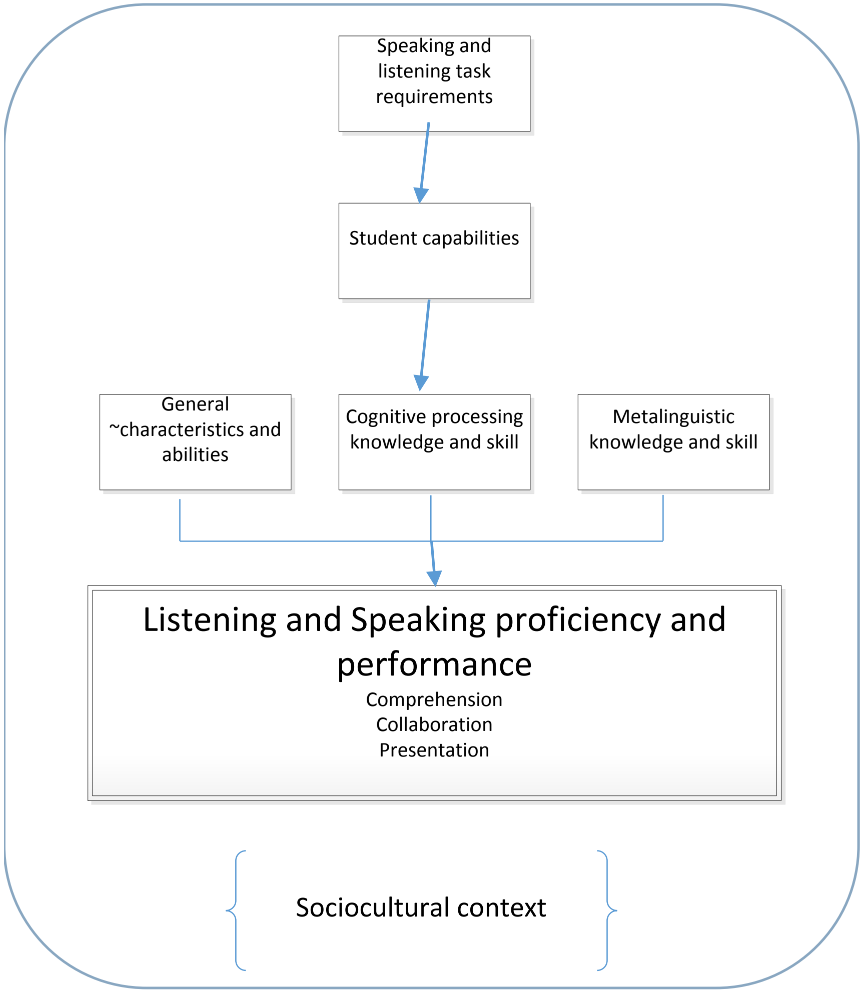
- "Diagnosing" stage in a learning progression can be summative or formative, given a systemic view
- Limitation is in collecting learner responses (technology), not interpreting learner performance (learning progression)
- Item performance and item statistics
- Difficulty
  - Area performance task=low difficulty
  - Area of the swimming pool=moderate difficulty
- Discrimination
  - Area performance task=low discrimination
  - Area of the swimming pool=high discrimination
- Informative
  - Area performance task=high information
  - Area of the swimming pool=low information

Illustration: Common Score Speaking and Listening (CCSL) learning and practice system

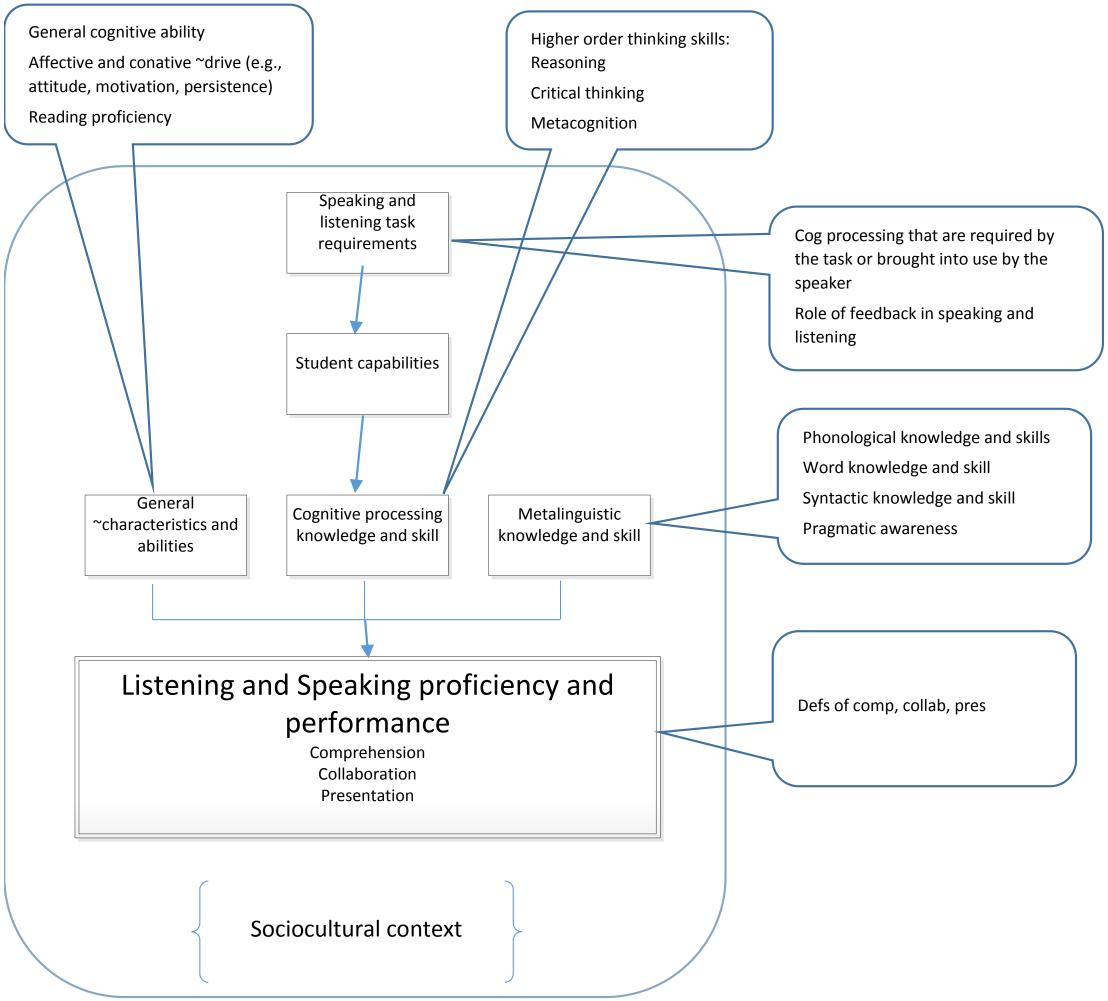
Prototype learning and formative feedback task: *Year Round School*



Model of learning and performance in speaking and listening



Model (cont.)


General cognitive ability
Affective and conative ~drive (e.g., attitude, motivation, persistence)
Reading proficiency
Higher order thinking skills:
Reasoning
Critical thinking
Metacognition
Speaking and listening task requirements
Cog processing that are required by the task or brought into use by the speaker
Role of feedback in speaking and listening
Student capabilities
Phonological knowledge and skills
Word knowledge and skill
Syntactic knowledge and skill
Pragmatic awareness
General ~characteristics and abilities
Cognitive processing knowledge and skill
Metalinguistic knowledge and skill
Listening and Speaking proficiency and performance
Comprehension
Collaboration
Presentation
Defs of comp, collab, pres
Sociocultural context

Evidence table

Knowledge, Skill	Activity	Evidence
1c. Pose questions that connect the ideas of several speakers and respond to others' questions and comments with relevant evidence, observations, and ideas.	**Stage 1.** Examinee views three videos, poses two questions	**Score Point 2**: Examinee poses two questions that connect the ideas of the three speakers, citing evidence from the three speakers.
4. Present claims and findings… **6.** Adapt speech to a variety of contexts and tasks…	**Stage 4**. Examinee makes oral presentation to classmates	**Score Point 3**: Examinee successfully and convincingly presents claims and findings by…

Thanks!

steve.ferrara@pearson.com

paul.Nichols@pearson.com

kristen.dicerbo@pearson.com

emily.lai@pearson.com

Research & Innovation NETWORK

References

Abt, C. (1970). *Serious games*. Lanham, MD: University Press of America.

Behrens, J. T., DiCerbo, K. E., and Ferrara, S. (2012). *Intended and unintended deceptions in the use of simulations*. Paper presented in the invitational research symposium on Technology Enhanced Assessments, Washington, DC. (See http://www.k12center.org/events/research_meetings/tea.html)

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Educational Technology & Society*, *17*(1), 17–28.

DiCerbo, K. E., & Behrens, J. T. (2014). *Impacts of the digital ocean*. London: Pearson. (Available from https://research.pearson.com/content/plc/prkc/uk/open-ideas/en/articles/a-tidal-wave-of-data/_jcr_content/par/articledownloadcompo/file.res/3897.Digital_Ocean_web.pdf)

Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A., Hao, J., Corrigan, S., et al. (2014). *Psychometric considerations in game-based assessment*. Redwood City, CA: GlassLab. (Available from http://glasslabgames.org/research/)

References (cont.)

National Research Council. (2001) *Knowing what students know: The science and design of educational assessment*. J. Pellegrino, N. Glaser (Eds.). Washington, DC: National Academy Press.

Nichols, P. D., Ferrara, S., & Lai, E. (2014 in press). Principled Design for Efficacy: Design and development for the next generation of assessments. In R. Lissitz and H. Jiao (Eds.), *The next generation of testing: Common core standards, Smarter Balanced, PARCC, and the nationwide testing movement*. Baltimore: Paul Brookes Publishing.

Nichols, P., Lai, E., Steedle, J., DiCerbo, K., & Ferrara, S. (2014 in press). Estimating reliability for learning analytics in games. In ***

Chudowsky, & R. Quellmalz, E., Timms, M., & Schneider, S. (2009). *Assessment of student learning in science simulations and games*. Washington, DC: National Academies Board on Science Education.

Salen, K., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. Cambridge, MA: MIT Press.

References (cont.)

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.) *Computer games and instruction.* Charlotte, NC: Information Age Publishers.

Stiggins, R. J. From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan*, *83*(1), 758-765.