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
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
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
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

Find evidence

Evaluate evidence

Critique argument

Battle robot

Equip robot

Construct argument
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

Game

Classroom Activity

Learning Progression

Professional Development

Performance Assessment

Student Profile
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

**Length Progression**
- Kindergarten: Unintentional view of measurable attributes
- Grade 1: Intentional awareness of measurable attributes
- Grade 2: Endpoint alignment
- Grade 3: Consistency of length

**Area Progression**
- Grade 1: Perceptual coordination of attributes across figures
- Grade 2: Internalization of informal area unit benchmarks
- Grade 3: Attribute of area

**Figure Composition/Decomposition Progression**
- Grade 1: Internal understanding of conservation of area
- Grade 2: Internalization of a formal formula for area
- Grade 3: Visualization of 2D shapes as a collection of area composites (units of units)

**Geometric Shapes Progression**
- Grade 1: Undifferentiated view
- Grade 2: Internalized informal formulas for perimeter
- Grade 3: Adopt formal formula for area

**Axis of Learning Progression**
- Kindergarten: Undifferentiated view of measurable attributes
- Grade 1: Endpoint alignment
- Grade 2: Consistency of length
- Grade 3: Attribute of area
- Grade 4: Area is additive
- Grade 5: Distributive property
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**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Source(s)</th>
<th>1st Step</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figures are drawn to scale</td>
<td>Joram, E. G., A. J., Bertheau, M., Gelman, R., &amp; Subrahmanyam, K. (2003). Children’s use of the reference point strategy for measurement estimation. Journal for Research in Mathematics Education, 36(1), 4 - 23.</td>
<td>Area: Internalized, formal area unit</td>
<td>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</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Length: Internalization of a mental ruler</td>
<td></td>
</tr>
<tr>
<td>Shapes are physically manipulable</td>
<td>Clements, D. (2011). Geometric and spatial thinking in early childhood education. In Clements &amp; Sarama (Eds.) Engaging young children in mathematics. Mahwah, NJ: Lawrence Erlbaum Associates.</td>
<td>Area: Perceptual coordination of attributes</td>
<td>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</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Length: Spatial structuring of length, length-unit/iteration</td>
<td></td>
</tr>
<tr>
<td>Level of scaffolding provided to help visualize internal structures of 2D shapes(e.g., presence/type of grid)</td>
<td>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.</td>
<td>Area: Visualization of 2D shapes as collections of area units, collections of area composites, or multi-level collections of area composites</td>
<td>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</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexible tool use (representing both move and less sophisticated strategies)</td>
<td>PATL project hypothesis</td>
<td>Area: Visualization of 2D shapes as collections of area units, collections of area composites, or multi-level collections of area composites, abstract informal formula for area</td>
<td>Student uses units to spatially structure both 1D and 2D spaces, using unit iteration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Geometry: Single-level, two-level, and multi-level composition</td>
<td>Students uses 1D and 2D structure to reason about area or length</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Length: Spatial structuring of length, length-unit/iteration, concepts &amp; skills of ruler use</td>
<td>Students uses a ruler to measure length of objects and calculate perimeter and area</td>
</tr>
<tr>
<td>Shape attributes(e.g., regular vs. irregular, rectangular vs. non-rectangular, conventional vs. unconventional)</td>
<td>Clements, Wilson, &amp; Sarama, 2004</td>
<td>Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition, visual/syncretic, descriptive, analytic, abstract</td>
<td>Students recognize a wide variety of shapes by their attributes or properties and identify and describe those attributes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Length: Spatial structuring of length, length-unit/iteration, concepts &amp; skills of ruler use</td>
<td>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</td>
</tr>
<tr>
<td>Shape dimensions (keep side lengths within 10 to minimize routine counting, addition, or multiplication errors)</td>
<td>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.</td>
<td>Area: Abstract informal area formula, adopt formal area formula</td>
<td>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</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Length: Abstract informal perimeter formula, adopt formal perimeter formula, length is additive, fluency in calculating perimeter</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Length: Spatial structuring of length, length-unit/iteration, concepts &amp; skills of ruler use</td>
<td>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</td>
</tr>
<tr>
<td>Shape dimensions given or not</td>
<td></td>
<td>Area: Adopt formal area formula</td>
<td>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.</td>
</tr>
<tr>
<td>Feature</td>
<td>Source(s)</td>
<td>LP Stage</td>
<td>Evidence</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
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 |
Geometry: Composition by trial and error, composition by attributes, single-level, two-level, and multi-level composition  
Length: Spatial structuring of | 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

Giraffe House

This is the giraffe house. How many tiles would it take to completely cover the giraffe house?

Number of Tiles: 0
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

<table>
<thead>
<tr>
<th>Activity 10 Title: Giraffe House</th>
<th>This is the giraffe house. How many tiles would it take to completely cover the giraffe house?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief Description of the Activity (this should indicate how this activity ties into the overall scenario or narrative of the PT)</td>
<td></td>
</tr>
<tr>
<td>Strand(s) of the Learning Progression that the Activity Targets</td>
<td>Area</td>
</tr>
<tr>
<td></td>
<td>Figure composition and decomposition</td>
</tr>
<tr>
<td>Stage(s) of the Learning Progression that the Activity Targets</td>
<td>Figure composition and decomposition: Single-level composition</td>
</tr>
<tr>
<td></td>
<td>Area: Using area units to measure area, Area unit iteration</td>
</tr>
<tr>
<td>Grade Level(s) that the Activity Targets</td>
<td>Grades 1-2</td>
</tr>
<tr>
<td>CCSS for Math</td>
<td>2.G.A.2 Partition a rectangle into rows and columns of same-size squares and count to find the total number of them.</td>
</tr>
</tbody>
</table>
### Stage 3: Construct reusable tools and templates

<table>
<thead>
<tr>
<th>Evidence to collect (observable behaviors or potential student responses)</th>
<th>Single-level composition: Student iterates individual unit squares to structure the shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area unit iteration: In iterating individual unit squares, student does not leave gaps in between unit tiles</td>
<td></td>
</tr>
<tr>
<td>Using area units to measure area: The student confirms the number of unit squares covering the bottom of the shape</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Content features</th>
<th>Shapes are physically manipulable, which allows students to compose or decompose composite figures and measure by iterating area units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape dimensions: Side lengths within 1-10 to minimize routine counting, addition, or multiplication errors</td>
<td></td>
</tr>
<tr>
<td>Shape attributes: Regular rectangle so area formula can be applied</td>
<td></td>
</tr>
<tr>
<td>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</td>
<td></td>
</tr>
</tbody>
</table>
## Stage 3: Construct reusable tools and templates

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

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

60 feet

20 feet
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

- Speaking and listening task requirements
- Student capabilities
- General characteristics and abilities
- Cognitive processing knowledge and skill
- Metalinguistic knowledge and skill

Listening and Speaking proficiency and performance
- Comprehension
- Collaboration
- Presentation

Sociocultural context
Model (cont.)

Listening and Speaking proficiency and performance

- Comprehension
- Collaboration
- Presentation

General characteristics and abilities
- General cognitive ability
  - Affective and conative “drive” (e.g., attitude, motivation, persistence)
  - Reading proficiency

Cognitive processing
- Knowledge and skill

Student capabilities

Speaking and listening task requirements

Higher order thinking skills:
- Reasoning
- Critical thinking
- Metacognition

Cog processing that are required by the task or brought into use by the speaker
- Role of feedback in speaking and listening

Metalinguistic knowledge and skill
- Phonological knowledge and skills
- Word knowledge and skills
- Syntactic knowledge and skills
- Pragmatic awareness

Def of comp, collab, pres

Sociocultural context

- General cognitive ability
- Affective and conative “drive” (e.g., attitude, motivation, persistence)
- Reading proficiency

Higher order thinking skills:
- Reasoning
- Critical thinking
- Metacognition

Cog processing that are required by the task or brought into use by the speaker
- Role of feedback in speaking and listening

Metalinguistic knowledge and skill
- Phonological knowledge and skills
- Word knowledge and skills
- Syntactic knowledge and skills
- Pragmatic awareness

Def of comp, collab, pres

Sociocultural context
### Evidence table

<table>
<thead>
<tr>
<th>Knowledge, Skill</th>
<th>Activity</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1c.</strong> Pose questions that connect the ideas of several speakers and respond to others’ questions and comments with relevant evidence, observations, and ideas.</td>
<td><strong>Stage 1.</strong> Examinee views three videos, poses two questions</td>
<td><strong>Score Point 2:</strong> Examinee poses two questions that connect the ideas of the three speakers, citing evidence from the three speakers.</td>
</tr>
<tr>
<td><strong>4.</strong> Present claims and findings...</td>
<td><strong>Stage 4.</strong> Examinee makes oral presentation to classmates</td>
<td><strong>Score Point 3:</strong> Examinee successfully and convincingly presents claims and findings by...</td>
</tr>
</tbody>
</table>
Thanks!

steve.ferrara@pearson.com

paul.Nichols@pearson.com

kristen.dicerbo@pearson.com

emily.lai@pearson.com
References


References (cont.)


References (cont.)

