Assessing the Common Core State Standards in Reading With *The Sports Network 2*

Jane Canner, Mary Schearer, Nicole Guagliardi, Classroom, Inc.
Email: jcanner@classroominc.org, mschearer@classroominc.org, nguagliardi@classroominc.org

Abstract: Classroom, Inc. (CI) developed and piloted an 8th grade learning game targeting the Common Core State Standards (CCSS) in Reading. CI built on 20 years experience using workplace simulations to teach literacy, and developed *The Sports Network 2 (TSN-2)*, leveraging advances in gaming and assessment. *TSN-2* puts students in the role of managing director of a sports media company, giving them reading and workplace problems to solve. Piloted in 2012 with over 400 New York City and Chicago students and their teachers, teachers agreed that the game was strongly connected to the CCSS, and students reported being engaged and learning from it. Assessments on the CCSS from within the game showed modest student performance; scores were strongly correlated with scores from a standardized reading test, the *Measures of Academic Progress.* This research provides early validation for an approach within a game to assess reading skills based on CCSS.

Introduction

Classroom, Inc. (CI), a nonprofit educational organization that has used virtual workplace simulations to improve adolescents' literacy, critical thinking and career readiness skills for over 20 years, has served over 700,000 students and 10,000 teachers across the country. CI's digital literacy programs in public, charter, and parochial schools are most often used in middle and high school summer and after school programs, and in classes that include a substantial number of struggling readers.

In 2011, a Next Generation Learning Challenges (NGLC) grant enabled CI to develop and pilot a prototype for a new program that would leverage advances in game technology, real-time progress reporting, and differentiated learning pathways to help low-achieving 8th graders master the rigorous new Common Core State Standards (CCSS) in Reading, while connecting them to the world of work.

This paper briefly describes the game, its reading demands, and the pilot; and provides preliminary validity data on *TSN-2* as a reading measure. Because most digital learning games are now focused on math and science rather than reading (Schwartz, 2013), this work is significant in showing the potential for developing games to address the complex reading demands of the new CCSS in actual classrooms.

The Game

Games have the potential to be "good assessment engines" (Shaffer & Gee, 2012) and to link standards, instruction, and assessment (Phillips & Popovic, 2012; Shute, 2011), but games that actually do this, particularly for the new CCSS, in ways that schools can use are few and far between. CI's challenge was to develop and pilot test an engaging game for adolescents that could both help improve struggling students' CCSS-related reading performance in school, and provide valid assessment information.

The program developed, *The Sports Network 2*, is a simulation that puts students in the role of managing director of a sports media company and presents them with a rich array of reading and workplace experiences. The game takes about 20 class periods to play—15 online and 5 offline. The continuous narrative includes five game quests, each representing a "day at work". The game addresses the following four Grade 8 Common Core State Standards (CCSS) for reading informational text:

Key Ideas and Details

- RI.8.1. Cite the textual evidence that most strongly supports an analysis of what the text says explicitly as well as inferences drawn from the text.
- RI.8.2. Determine a central idea of a text and analyze its development over the course of the text, including its relationship to supporting ideas; provide an objective summary of the text.

- RI.8.3. Analyze how a text makes connections among and distinctions between individuals, ideas, or events (e.g., through comparisons, analogies, or categories).

Craft and Structure

- RI.8.4. Determine the meaning of words and phrases as they are used in a text, including figurative, connotative, and technical meanings; analyze the impact of specific word choices on meaning and tone, including analogies or allusions to other texts.

At the outset, students learn that TSN-2 is swiftly losing its teen audience; the managing director must find a way to draw large numbers of teens to TSN-2 quickly. All five quests lead to that objective. This is a high-stakes proposition because people's jobs hang in the balance. The five quests are:

> *How Do We Draw Teens Back to TSN-2?* Solve TSN-2's dwindling teen viewership problem.
> *What Is Our Audience? What Is Our Topic?* Determine exact target audience and topic for new TSN-2 program aimed at teens.
> *What Sport Should Our Pilot Feature?* Scout locations/choose a setting for "Teens & City Sports."
> *How Can We Pitch Our Pilot?* Gather information, interview site coordinator, and create a storyboard for a pilot of "Teens & City Sports."
> *Will Our Pilot Get the Green Light?* Complete and present a pitch for a pilot to TSN-2's president.

In each quest, students complete several main reading activities, where they read varied kinds of informational text, including typical workplace communications such as contracts and emails, and other materials, e.g., opinion pieces, articles, conversations, interview transcripts, and research reports. These "main" texts are each 1-3 pages in length, and are at the low-to-mid 8th grade reading level (average Lexile of 1010). For some activities, students are routed seamlessly to easier (6th grade reading level) or more challenging (mid-high 8th grade reading level) activities, based on their initial performance.

Embedded Assessment Approach

Evidence-centered design (Shute, 2011) requires game developers to ask:

- What do we want to say about the student?
- What observations would provide the best evidence for what we want to say?
- What tasks will enable us to make these observations?

Classroom, Inc. addressed these questions early in the game development process.

What do we want to say about the student?

Classroom, Inc. adopted its overall goal from the CCSS documents: Students will be prepared for college and careers by being able to read and comprehend rigorous, grade-level, nonfiction text; make connections among ideas in the texts; and consider a wide range of textual evidence.

What observations would provide the best evidence for what we want to say?

When presented with a variety of text types and formats in a game, students will show that they comprehend the text by answering questions correctly about what they have read.

What tasks will allow us to make these observations?

The tasks needed include giving students nonfiction text passages of varied types at the 8th grade level (as measured by Lexiles and CCSS complexity levels) within game mechanics that require them to read these texts and demonstrate their understanding before proceeding within the game.

CI and its game development partner, Filament Games, focused most early efforts on developing appropriate tasks in the form of game mechanics to elicit these observations. To further understand and operationalize the CCSS, CI carefully reviewed the CCSS documents, including sample texts, and recommended ways to ensure text complexity; reviewed the "Publishers' Criteria for the CCSS in ELA and Literacy" written by CCSS authors; consulted with experts in assessment and standards; and developed our own internal guidelines for text types to use and behaviors to elicit. CI then developed descriptions of mechanics to use, and guidelines for nonfiction text writers. To maintain workplace

authenticity and narrative flow, CI used those mechanics throughout the game, often letting the narrative suggest placement, and mixing them up to keep students engaged. CI ensured ample coverage of each standard throughout the game, to generate sufficient information to assess students on each standard.

Because of competing requirements, *TSN-2* ended up with a different number of activities for each CCSS and a different number of "items" per activity, but enough to generate assessment data. (The game included 29 main reading activities/assessments with 260-265 items in all. The number of activities per standard ranged from 4 to 14; and the number of items per standard ranged from 50 to 87. See Table 1 for more specific breakdown.)

Sample Game Mechanics/Assessments

CI decided on nine game mechanics targeted to specific CCSS to use throughout the game, two of which--Idea Centralizer and Sorting Organizer—are featured in this paper. *Eight of these mechanics—not the open-ended Writing activity—served as both instructional tasks and embedded assessments.*

Piloting *TSN-2* provided valuable insight into how well the mechanics worked and why. In many cases, CI adapted these mechanics from existing ones used by Filament Games, to accommodate our aggressive funding and publishing schedule. A description and brief analysis of *two* oft-used mechanics follows.

Idea Centralizer (for CCSS RI 8.1 and 8.2). Students use this mechanic both to help them examine details that illustrate, explain, or support a main idea and to analyze the meaning of text within a larger context. For example, students read several ideas for TV shows. Knowing TSN-2's goals for creating a TV show that will be popular with teens—and profitable for TSN-2—students use the mechanic to determine the elements of proposed program ideas that do and do not meet TSN-2's goals.

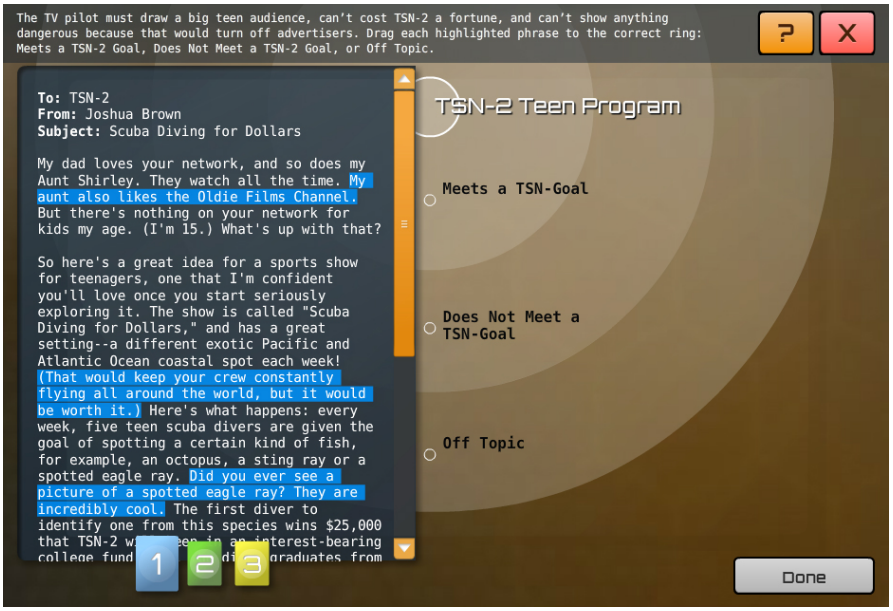

Figure 1: **Sample *TSN-2* Idea Centralizer Mechanic**

Students could have played up to seven *Idea Centralizer* activities in all. Over 300 students played an average of 4.5 *Idea Centralizers*, and answered 58% of the items correctly. This mechanic works best with clear criteria for text analysis. It has a straightforward interface that enables students to concentrate on the text. Its' tabs accommodate longer texts, and is an effective mechanic for its specific use.

Sorting Organizer (for CCSS RI 8.3).Students compare details of three different people, places, or ideas. They must read and make judgments about text. In the following example, students compare the settings

of three urban sports programs to determine which would make the best to feature in a pilot episode by determining which features of each setting meet the criteria of making a strong pilot TV show for TSN-2.

***Figure 2:* Sample *TSN-2* Sorting Organizer Mechanic**

Students could have played up to two *Sorting Organizer*s across two "days." Over 180 students played an average of 1.5 of these, and answered 63% of the items correctly. This mechanic works well because there is no possibility of giving away an answer in target text fragments and the activity does not call on students to focus on some details to the exclusion of others. Students had to carefully read text, think about it from different angles, and—when the mechanic interaction was complete—synthesize all the information to make comparisons. It was not, however, designed for sufficiently lengthy text.

Research Methods

CI worked with the NGLC Wave 2 evaluator, SRI International, to collect multiple types of outcomes and implementation data to assess program efficacy. CI sought to address the following questions related to student reading performance, this paper's focus:

- What are the challenges in designing embedded assessments, particularly for 8th grade CCSS for reading? This was relatively uncharted territory; learnings could help others seeking to instruct students and assess learning toward CCSS using technology.
- How do students perform on the four CCSS on embedded assessments, and how does this correlate with a standardized reading test, *Measures of Academic Progress Reading (MAP)*?

Participants

Sixteen teachers and over 400 students in 17 classes in eight NYC and Chicago public schools participated. Most teachers taught English Language Arts (ELA); others taught computers, electives, or ESL. Most used the program in regular classrooms with laptop carts. Almost all students were in 8th or 9th grade (96%); just over one-half were girls (55%); and the sample was ethnically diverse, with 53% identifying themselves as Black, 22% Latino, 13% Asian, and 11% White. Thirty-five students with special needs (8.4%) participated.

Measures and Data Collection

CI worked with SRI to collect multiple types of data. SRI collected standardized reading test data (pre and post-program data from an online assessment, the *MAP,* and an end-of-program teacher survey. CI

collected brief pre- and post-program student surveys on deeper learning competencies and student engagement, and gathered CCSS student performance data through *TSN-2*'s embedded assessments.

Embedded Assessment Data

Because the use of embedded assessment data is evolving, additional information on CI's approach and use of these data is provided. (Note that since the main game activities students played served as the assessments, the terms activity and assessment are used interchangeably here.) Students' responses within the game to the mechanics/embedded assessments were automatically captured and allowed CI and teachers to better understand students' reading competency. CI captured a wide array of data, including time spent in each activity, performance on all items in each activity, and response accuracy.

CI initially reviewed responses by mechanic and found that some worked better than others, and is now using those findings for new literacy games being developed. CI then examined the volume and accuracy of student responses around each CCSS.

For all but open-ended written items, responses were scored for accuracy, and then a simple *percentage correct* was calculated for all items within each activity. These percentage-correct scores were then averaged for each standard. This approach gave the same weight to each activity regardless of the number of items, and was akin to scores in a teacher's grade book, where a report card grade is an average of scores on multiple assessments, each with different numbers of items. This modest "grade book" strategy was used to calculate overall scores, and scores for each activity and each CCSS.

Results

While Classroom, Inc. collected a variety of data, this paper focuses primarily on reading outcomes, especially those from *TSN-2*'s embedded assessments. Highlights of survey results focusing on reading and learning are provided first, followed by embedded assessment results.

Highlights of Survey Results

Teachers agreed that *TSN-2* was a valuable and engaging educational experience for students, and that it had *a strong connection with the Common Core State Standards* (78%), and met *the learning needs of their students (67%)."* Two-thirds of teachers reported that learning gains were greater with *TSN-2* as compared to a traditional class. Teachers told us that *TSN-2* was a learning experience for them, particularly regarding the CCSS.

> *"I've never had anything that addressed CCSS like this!"*
> *"TSN-2 hugely addressed CCSS, particularly in vocabulary."*
> *"This game addressed the CCSS more than anything else I am doing."*

Students were highly engaged in the game, and learned by playing it, according to their survey responses. Large majorities of students agreed with the following statements:

> *Using the computer made learning more fun. (90%)*
> *I learned things that UI can use when I grow up. (90%)*
> *I am proud of my work in TSN-2. (88%)*
> *What I learned will help me do better in school. (88%)*
> *I am smarter than I thought I was. (86%)*
> *TSN-2 made me want to learn other things. (81%)*

Highlights of Embedded Assessment Results

This paper presents and discusses two sets of results from *TSN-2*'s embedded assessments: student performance on the targeted CCSS reading standards, and the correlation of the embedded assessment reading performance with performance measured by the standardized *MAP* Reading tests.

As Table 1 shows, struggling 8th grade readers who played *TSN-2* performed reasonably well overall on most standards, with percentages correct ranging from 43%-59% and averaging 53%. The texts were written at a rigorous low-mid 8th grade CCSS level. (Please note that the poorest performance—43% on

Standard 8.3—was an anomaly and was largely due to one problematic mechanic that we are revising. If this activity type is ignored, the average percentage correct for Standard 8.3 increases to 63%.)

Main Activities	Number of Activities	Number of Items	Average Number of Students Who Played Each Activity	Average Percentage Correct
Standard 8.1 Cite the textual evidence . . .	14	57-62	196	52%
Standard 8.2 Determine a central idea . . .	6	87	127	59%
Standard 8.3 Analyze how a text makes connections . . .	4	66	178	43%
Standard 8.4 Determine the meaning of words and phrases . . .	5	50	168	55%
Overall	**29**	**260-265**	**174**	**53%**

***Table 1*: *TSN-2* Performance on Embedded Assessments by Standard**

Table 2 presents correlations between the *TSN-2* embedded assessment and the post-MAP Reading scores. *The correlations are moderate to strong, showing that TSN-2 is likely measuring the same traits as the MAP, and that this game does indeed assess students' reading comprehension. These data provide preliminary validation of this new game as an assessment of reading comprehension.* (Please note that correlations between embedded assessments and the MAP Reading pre-test were also performed, and were similarly high and also all significant at the 0.01 level.)

TSN-2 Embedded Assessment Scores		MAP Posttest Total Score
ALL Main Score	r	.685**
	sig	.000
	n	106
8.1 Main Score	r	.564**
	sig	.000
	n	106
8.2 Main Score	r	.602**
	sig	.000
	n	102
8.3 Main Score	r	.352**
	sig	.000
	n	96
8.4 Main Score	r	.653**
	sig	.000
	n	101

** Correlation is significant at the 0.01 level.

***Table 2*: Correlations between Embedded Assessment Scores and MAP Post Test Scores: Pearson Two-Tailed Correlation**

Analysis and Discussion

The focus of our analyses to-date has been to understand the data we were able to capture, and—most importantly—what those data say about students' reading performance. *TSN-2* by design, did not include a formal pre- and post-test. Although information on student performance was gathered throughout the game, with this short program exposure (20 hours), CI did not assess gains, but rather performance. New longer programs CI is developing will have this capacity.

Regarding student performance, along with embedded assessment data, we have teacher and student survey feedback, class observations and anecdotal evidence indicating that students were reading more and better, taking books out of the library, etc., after their *TSN-2* experience. We targeted struggling eighth graders and found those students in general getting 50-60% of the items related to comprehending 8th grade level text correct, performing moderately well with difficult material. Feedback from students and teachers indicated that students, even those with special needs, performed reasonably well with the grade 8 informational text, and were motivated and engaged.

CI developed a learning game, not a standardized assessment. CI focused on engaging and challenging students with authentic workplace situations, and—at the same time—instructing and assessing those students on rigorous 8th grade standards. The challenge was to engage students and have them move through the game without disrupting the flow with external assessments. Given the time and other limits, CI gathered considerable data useful to CI, teachers, and the broader learning community in the pilot.

Evidence from the *TSN-2* pilot indicates that it engaged teachers and students in CCSS-appropriate literacy activities, and that students were able to perform moderately well on rigorous nonfiction text within this game setting. Key remaining questions are how much assessment data game developers must collect and provide to have their games used both widely and well in real schools on a regular basis, and also be valid and reliable measures of student performance on the CCSS.

References

Phillips, V. & Popović, Z. (2012). More than child's play: Games have potential learning and assessment tools. *Kappan Magazine*, *94 (2)*, pp. 26-30.

Schwartz, K. (2013). What online tools work for teaching language arts? Retrieved from http://blogs.kqed.org/mindshift/2013/01/what-online-tools-work-for-language-arts.

Shaffer, D.W. & Gee, J. P. (2012). The right kind of GATE: Computer games and the future of assessment. *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern Research: Information Age Publications*.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction*, pp. 503-523.