Creating Games as Reader Response and Comprehension Assessment

Constance Ulmer, Mary Timothy, Lynne Bercaw, Susan Gilbert, Jody Holleman and Mari Hunting

The definitions of “reading” and “reading comprehension” have moved beyond mere explicit recall, word recognition, and mastery of phonemic decoding. The definitions now include—or perhaps are even replaced by—the dynamic, reciprocal interactions among reader, text, and the context of the reader’s prior literacy schema. No longer is the quiet, private, solipsistic model of the reading process adequate; the new model, rather, is that reading is an interactive and complex process. The process through which the dynamic interaction of the reader’s background knowledge, the information inferred by the written language, and the reading situation context (Dutcher, 1990) is constructing meaning. This new, more divergent definition of reading requires that it be accompanied by what has come to be known as authentic assessment. Such a new assessment model should be characterized by the following: 1) It should address the reader’s cognitive ability to construct meaning out of what is implied in the text, 2) it should assist in developing reading fluency, skills, and strategies, and 3) it should honor both the reader’s literacy context and ability to make cognitive and affective leaps based on a synthesis of old and new information (Wiggins, 1990).

Authentic assessment may be best defined as the “direct examination of student performance on worthy, intellectual tasks” (Wiggins, 1990). In light of this definition, whose key word is “worthy,” the traditional, empirically-based formal skills-testing used for reading comprehension ceases to be a valid instrument for measuring so complex and
robust a process as is currently recognized to be the case. Further, O’Neal (1991) notes that the traditional, “imperfect” reading comprehension testing instruments are misleading in that they offer little in the way of assessing inferential, critical thinking, or of the reader’s affective responses.

Because the reading process cannot be directly observed, Powell (1989) stated, “all scores or data produced by tests of reading are indirect measures of the reading process” (p.1). The challenge, then, is to create authentic, holistic reading comprehension assessments, a task which has historically been in the purview of researchers, reading experts, cognitive psychologists, and educators. It is teachers in particular who are taking a more assertive role in creating authentic evaluation methods that, in turn, address the particular needs of their students (Shapiro & Kilbey, 1990; Stern & Shavelson, 1981). Given the greater individual teacher role, students have greater opportunity to demonstrate what they have actually learned (as opposed to their ability to fit imposed descriptors), while teachers can monitor and adjust their instruction accordingly.

A recurring criticism of authentic assessment is aimed at its subjectivity, calling into question its validity and reliability. Wiggins (1990) responds to this criticism this way:

Genuine accountability does not avoid human judgment. Training sessions, model performances used as exemplars, audit and oversight policies, as well as such basic procedures as having disinterested judges review students’ work “blind” occur routinely throughout the professional, athletic, and artistic worlds in the judging of performance. (p.4)
Extrapolating such accepted practices to the realm of reading and reading comprehension should require no extraordinary leap of faith, despite the common criticism of “subjectivity.”

Reader Response Theory

Of the enormous collection of alternatives to the assessment of reading comprehension, reader response theory has made a profound impact on the vast corpus of assessment tools in the past three decades. Based on the literary theory of Rosenblatt (1995), reader response theory emphasizes a reciprocal relationship between the reader and the text. The reader brings his or her cultural background and socialization to the text. The meaning, therefore, that the reader creates from the text is synthesized with the text and the life context the reader brings to text (Rosenblatt, 1991). This differs in contrast to the “communication model” where the reader attempts to decipher the message conveyed by the author (Many & Wiseman, 1992).

The reader response theory first emerged with Rosenblatt’s literary theory in Literature as Exploration, (first edition 1938). Rosenblatt emphasized not only the transaction theory between reader and text as a way to view the reading process, but how reader response theory influences the teaching of reading and literature (Many & Wiseman, 1992). Rosenblatt describes two stances toward reading: efferent and aesthetic (Rosenblatt, 1991). Rosenblatt (1991) found that a reader takes an efferent stance when the goal of reading is to gain information from the text (e.g. textbooks, newspapers, reference materials, etc.). She and others contend that when responding to text, an aesthetic stance is appropriate where the reader participates in the experience created through the personal transaction with the text (Many & Wiseman, 1992; Rosenblatt,
Further, the aesthetic stance results in the personal, experiential aspects of meaning.

Various studies support Rosenblatt’s emphasis on the aesthetic response to literature (Cox & Many, 1992; Hickman, 1983; Many, 1990). Many (1990) studied upper elementary grades and middle graders’ responses to literature and found that the level of understanding when students’ responded aesthetically were significantly higher than when they responded with efferent responses.

Games As a Reader Response

In *The Significance of Rosenblatt on the Field of Teaching Literature*, Church (1997) states “…during any one reading experience readers may shift back and forth along a continuum between efferent and aesthetic modes of the reading processing” (p. 72). It is often during the shifting (transaction) from these two stances that readers began to acquire knowledge. Thus assessment of reading comprehension becomes quite complex and both stances must be inclusive in the assessment process.

Alternative assessment studies such as retellings, have introduced the changes in research to consider both stances (Irwin, & Mitchell, 1983, Feathers, & White, 1985, Ulmer, 1992.). Reader response theories indicate that both a communicative and a personal response approach to reading is the intention of the tasks assigned to learners (Hirvela, 1996), making the learners active participants in the reading process. The interactions that occur during the reading process involve both the efferent and aesthetic modes of learning. Assessing reader responses through activities that help activate learning is imperative if evaluating both modes of reader response theory (efferent and aesthetic), which includes all three parameters of the reading process (textual, affective,
Creating and playing games (as a form of reader responses activities) has been used to assess comprehension. They have often been used to assess the efferent modes of learning (textual differencing, recall of text, and memory). These studies considered the ‘text’ the focus of learning. (Curtner-Smith, 1996; Pride, Pride, Outman, & Iddings, 1999; Graham, & Williams, 2001). Other studies considered Vygotskian features of development important, relating the social cultural aspects with learning. These studies considered the “reader’ the focus of learning, the aesthetic modes of learning in their research.

Games promote active, student-centered learning, motivation, cooperation and social learning, and reduce risk of failure (Davis & Hollowell, 1977). Early studies using games rarely addressed content, but related games to the social aspects of the interactions occurring between players while playing. Some of the newer studies also follow this trend.

Brisk (1974) offered a seminar course at the University of Mexico where students created games during the entire seminar (various majors). Games were their only motivation and teaching strategy for the course. Students were to create games based on some concept presented in their discipline through lectures or reading. The instructions Brisk gave were “games should be entertaining, based on reality, relevant to a social situation and predictive – i.e., oriented the ‘the near future” (p. 91). Their colleagues had to be able to play the games during a one to two hour session. Even though Brisk did not evaluate the games thoroughly, he indicated that when students were engaged, they all seemed to be relating on different levels of the reading parameters (affective, cognitive,
and textual) (Feathers, 1981). Brisk (1974) concluded that constructing games were the stimulus students needed to do necessary work demanding their creative scholarship approaching serious deductive reasoning.

Much of the research on games is related to playing games or gaming (not creating games). Gaming is the manipulation of the games. These studies suggest that handling or playing the games enhances problem solving, critical thinking, recall of text, memory, and vocabulary, (Gaudart, 1999; Graham, & Williams, 2001). Fernie, (1988) believes, games such as checkers teach offensive and defensive alternatives that can intellectually motivate children. Most of these behaviors relate to textual and cognitive parameters of learning. Some of the newer studies using games focus on the construction of games by students as a learning tool. These address affective as well as textual and cognitive parameters in the learning process.

Playing games is only part of the learning process and comprehensive assessment is therefore difficult. Creating games allows for higher levels of thinking and more personal responses (affective parameter) to what is being learned. Domke (1991) and Spiegel (1990) found it important to create games for the students, but having students create their own games is equally important if learning is an active process. In creating games, children negotiate the rules in which they wish to play the games (King, 1986). Creating games teaches reasoning strategies and skills. Berrenberg (1991) states:

It is generally assumed that students learn and retain more when they are actively and personally involved with the course material… relatively few examination procedures incorporate this active approach. (p. 168)
Berrenberg says her students enjoyed creating the games and they said they learned more that way.

Research on games and gaming mostly focus on cognitive and textual engagements in the reading process, but there is still a need to further evaluate the affective engagements during the reading process. The aesthetic stance of reading connects the schema of processing (while reading) to the schema of producing (after reading). All aspects of reading (cognitive, affective, and textual) address both stances and; therefore, should be considered during assessment.

Methodology

Qualitative and quantitative research methods were chosen to explore the issues of creating games as an alternative assessment of reading comprehension in reader response. According to Bogden and Biklin (1998), "If you want to understand the way people think about their world and how those definitions are formed you need to get close to them, to hear them talk, and observe them in their day-to-day lives" (p.32). Qualitative allowed the researchers to approach the setting and participants without a predetermined hypothesis. Theory emerged from the data. This research focused on creating games as a reader response to assess reading comprehension.

Triangulation is an important part of qualitative research because it gives credibility to the data. Triangulation of sources is locating similar information from various sources and using similar analysis to confirm findings in the data. This was accomplished by having 5 raters cross-analyzing each of his or her findings.

Patton (1990) places the obligation on the researchers to be methodical in presenting sufficient details of data collection and the process used in analysis so others
can judge the quality of the product: validity is the credibility of the study and the reliability is the dependability of the research.

Prior to examining the games created by the 17 groups of preservice teachers, the 5 raters developed a Likert Scale for evaluating the games looking at textual, cognitive, and affective parameters. The 5 raters individually analyzed and rated each game looking at the three parameters. The raters then met as a group to compare findings. The validity of the study showed all of the games created produced the three parameters. The validity was shown when the individual ratings were compared and the numbers were consistent among the raters.

Participants

Ninety preservice teachers enrolled in their first educational course participated in this study. The preservice teachers represented a variety of majors and grade levels. The books used for the reading responses were children’s books.

Procedure

The students (preservice teachers), in a workshop prior to field placement, organized themselves into 17 groups. The workshops presented reading strategies, alternative assessments, and observations, followed by open discussions. The students were asked to divide themselves into groups of 5 or 6 and create a game related to one of the four books book presented in the workshops or one from four alternative books. Students were told to create a game related to the book. The one requirement was that the game had to be functional for children to play; therefore, directions were to be included. No other directions were given. Supplies were available for the students to use to create their games. When the games were completed, they were asked to write a brief reflection
about how they felt creating a game. Students were given one class period (50 minutes) to accomplish the assignment.

Data Analysis

Initial analysis showed in part what we expected: students’ textual responses to the literature were represented the most in the games while their cognitive responses were represented the least. The games were first submitted to qualitative content analysis then quantified for each reading parameter to produce tables. Five members of the research team individually examined each game through the lens of textual, affective, and cognitive parameters, using a holistic rubric based on the principles of a Likert scale (0-5 likert scale – 0 not indicated in the game, 5 highly present). For example in B10’s game rating, the research team’s individual scores in textual were 5,5,5, 5, and 5; affective were 0,5,1,0, and 3; and cognitive were 2, 5, 3, 2, and 4. The researchers rated the game highly textual and little affective. In B7’s game, the researchers agreed that the affective parameter was stronger. Those individual scores were: textual 3,3,2,3, and 2; cognitive 4,5,1,3, and 0; affective 3,4,3,4, and 4.

After individual ratings, the research team shared the scores (0-5) for each of the three parameters within each game. When there was a discrepancy between raters (e.g. 0 & 5), there was discussion of the rubric and the criterion upon which scores were based, until agreement was reached. Finally the scores were averaged for an overall for each parameter for each game. For example, Game B-10 received individual scores in the Cognitive Parameter of 2,5,3,2, and 4. These scores were averaged to give a 3.2 overall rating of Cognitive Parameter.

The total scores for each parameter were then analyzed. The Cognitive Parameter
score for each of the 17 games totaled 38.8 out of a total of 85 points (5 points possible for each of 17 games). The total for the Affective Parameter is 43.4, and the total for the Textual Parameter was 55.2.

We also assessed whether the games were predominantly class activities (didactic in nature) or if they were purely games (enjoyment without didactic components). Nine of the games were identified as class activities while 8 of the games were identified as “games”.

Results

The cognitive score for each of the 17 games totaled 38.8 (45%) out of a total of 85 points (5 points possible for each of the 17 games)(See Appendix A). The total for the affective parameter was 43.4 (51%) and the total for textual parameter was 55.2 (65%). Textual information rated the highest while cognitive was close behind affective. The researchers felt that affective rated higher than cognitive because students brought their own voices to the games they created. The games were also assessed to determine whether they were predominantly class activities (didactic in nature) or whether they were truly games (enjoyment without didactic components). Nine of the games were identified as class activities, while eight of the games were identified as games with winners and losers between players. The distinction between class activities and games indicates that preservice teachers did not fully understand the “gaming process.”

Discussion and Conclusion

The importance of these findings relate to what we teach in schools and how we evaluate what students are learning. Traditional assessment of reading comprehension has been empirically based on formal skills-testing, which can be misleading to teachers.
Having students create games as a form of assessment allows teachers to see how students are interacting with written text by what is demonstrated in the game. We have often heard it said that games are for children, but with the increase of computer and technology as “text” in the 21st century, the definition and evaluation of comprehension changes. Reading constantly involves interactions with text. Creating games lets students demonstrate what they learned from written text. By creating games as assessment tools, students (younger and older) are making meaning as they search and create questions and even find destinations to answer questions. This in turn creates new questions to ask.

Creating games goes beyond the who, what, when, and where. Students negotiate, evaluate, critique, plan, act, interpret, and postulate while reading and writing at the same time. Even though nine of the seventeen games were creative activities the other eight were games. In those games, as well as the activities, we did see all three parameters (textual, affective, cognitive) present, even though there were no significant differences, except textual being the highest.

Games were found to be useful as an assessment tool as well as an effective teaching strategy and learning instrument. The games indicated high levels of affective, cognitive, and textual parameters. Games encouraged collaboration, improved retention, promoted student interest, facilitated higher cognitive skills, allowed for effective problem solving, provided direct applications of knowledge, helped students synthesize information, and made students into decision makers.

When looking at cognitive (analyzing and synthesizing), affective (creativity and student voice), and textual (factual information) parameters found in reader responses, preservice teachers included textual information as evidenced in their games; however,
they varied in the cognitive and affective measures. Using games is a more authentic form of assessment of reading comprehension than other traditional forms of assessment. Basal text questions (based on traditional standardized test measures) are restricted in that they only ask certain, explicit information pertaining to the text. The non-traditional, play-based creation of games allows all levels of comprehension to be incorporated thus less restrictive on what is being learned.

From our findings we contend that there is value in creating games as a form of assessment for various levels of comprehension but very few teachers benefit from this knowledge. Why the hesitations to use games in more classrooms? A few conclusions could be drawn: 1. teachers are afraid to deviate from the traditional forms of assessment, 2. teachers’ expectations of what is to be learned must adhere to specific textual information, or 3) there is not enough time to have students create and play games and then develop a formula for evaluation. Any or all of the conclusions can be true for various reasons. Using games as evaluations or assessments in the classroom is radical and risky. Brisk (1974), Berrenberg, (1991), and Gaudart (1999) all agree, but they believe it is worth the risk, because the results were surprising and revealing as suggested in this study. Gaudart reacts to the risk of using games as a new technique in his conclusion. According to Gaudart (1999):

It is sad, therefore that techniques that can so effectively motivate learners are introduced into classes with a great deal of fear and trepidation. Simulation and games are techniques that have been sidelined for a very long time, and I have spent 20 years trying to get teachers to use the techniques in their classrooms. (p. 289)
He concludes that teachers teach the way they learned.

The authors in this study conclude that the games created by the preservice teachers indicated their higher level of comprehension through the inclusion of cognitive and affective parameters moving beyond textual information often measured. However, more education is needed to get teachers and preservice teachers to see the value of creating games in the classroom as assessment instruments and possibly teaching strategies.

References


Clearinghouse on Tests, Measurement, and Evaluation, Washington, DC.; American Institutes for Research, Washington, DC.


Paper presented at the Annual Conference of the International Reading Association, March, New Orleans, LA.


Reading Conference. Chicago, IL: National Reading Conference.


Table 1: Content Analysis for Protocols of Games

<table>
<thead>
<tr>
<th>Games</th>
<th>Cognitive</th>
<th>Affective</th>
<th>Textual</th>
<th>Games</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>4</td>
<td>3.6</td>
<td>2.4</td>
<td>X</td>
</tr>
<tr>
<td>B2</td>
<td>2.2</td>
<td>3</td>
<td>4.8</td>
<td>X</td>
</tr>
<tr>
<td>B3</td>
<td>1.2</td>
<td>.4</td>
<td>5</td>
<td>X</td>
</tr>
<tr>
<td>B4</td>
<td>2.2</td>
<td>2.4</td>
<td>4</td>
<td>X</td>
</tr>
<tr>
<td>B5</td>
<td>1.6</td>
<td>3</td>
<td>.8</td>
<td>X</td>
</tr>
<tr>
<td>B6</td>
<td>3.8</td>
<td>3.4</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>B7</td>
<td>2.6</td>
<td>3.6</td>
<td>2.6</td>
<td>X</td>
</tr>
<tr>
<td>B8</td>
<td>.6</td>
<td>1.4</td>
<td>2.2</td>
<td>X</td>
</tr>
<tr>
<td>B9</td>
<td>3</td>
<td>2.8</td>
<td>3.2</td>
<td>X</td>
</tr>
<tr>
<td>B10</td>
<td>3.2</td>
<td>1.8</td>
<td>5</td>
<td>X</td>
</tr>
<tr>
<td>B11</td>
<td>.8</td>
<td>2.6</td>
<td>1.6</td>
<td>X</td>
</tr>
<tr>
<td>B12</td>
<td>1.8</td>
<td>2.8</td>
<td>3</td>
<td>X</td>
</tr>
<tr>
<td>B13</td>
<td>3.8</td>
<td>2.8</td>
<td>3.6</td>
<td>X</td>
</tr>
<tr>
<td>B14</td>
<td>.8</td>
<td>2</td>
<td>4.4</td>
<td>X</td>
</tr>
<tr>
<td>B15</td>
<td>1.6</td>
<td>2</td>
<td>2.6</td>
<td>X</td>
</tr>
<tr>
<td>B16</td>
<td>4</td>
<td>3.4</td>
<td>3.4</td>
<td>X</td>
</tr>
<tr>
<td>B17</td>
<td>1.6</td>
<td>2.4</td>
<td>2</td>
<td>X</td>
</tr>
<tr>
<td>Total</td>
<td>38.8</td>
<td>43.4</td>
<td>55.2</td>
<td>8</td>
</tr>
<tr>
<td>Average</td>
<td>2.28</td>
<td>2.55</td>
<td>3.24</td>
<td>47%</td>
</tr>
</tbody>
</table>

** Ratings were based on 0-5 points. 5 containing highest indication of parameter.