Teachers' Uses of Students' Digital Annotations: Implications for the Formative Assessment of Reading Comprehension

Barry Brahier, Ph.D.
University of Minnesota, Center for Reading Research

August, 2006

Teachers' Uses of Students' Digital Annotations: Implications for the Formative

Assessment of Reading Comprehension

Barry Brahier, Ph.D.

University of Minnesota

Reading comprehension has been called the "bottom line" of reading instruction (McKenna & Stahl, 2003) and recent national reports call for improving the assessment of reading comprehension by addressing the complaint that most assessments are not useful to teachers (Snow, 2002; Snow & Biancarosa, 2003). Teachers' use of their students' annotations fits well with calls for literacy assessment reform. In arguing for additional assessment types, Tierney (1998) asserted that students' "learnings may be fleeting" (p. 376) and suggested teachers facilitate assessment by encouraging students "to keep traces of what they do" (p. 375). Winne and Hadwin (1998) argued specifically for teachers to examine students' margin notes as traces of their metacognitive effort, while Johnston (2003) urged teachers to notice and record the literate practices of their students. While examining students' annotations would be a step towards the implementation of these reforms, currently there are two significant barriers to the use of annotation in K-12 schools.

The first barrier to the use of annotation is the complete prohibition against marking in the textbooks students use in K-12 schools. This is for good reason as the common practice is that the school district, not the students, own the books. This barrier could be overcome by mandating that families buy textbooks for their children, or the

adoption of disposable workbooks, or both. However, these solutions would leave intact the second barrier: annotations on paper materials are not easily shareable between students and their teachers. Both of these barriers are being addressed through the increasing availability of digital annotation systems.

Digital annotation systems allow users to add marks and notes to existing digital text. In 2002, Wolfe described over 25 digital annotation systems, categorizing them as either in existence or development. While their features vary, they all overcome both of the barriers stated above by taking advantage of what Negroponte (1995) called "the differences between bits and atoms" (p. 11). Negroponte used the term *bits* to label items in the digital realm that have no physical existence (i.e., computer files such as those made by word processors, spreadsheets, and so on) and *atoms* to label anything that does have physical existence (i.e., paper, books, ink, pencil, etc.). The advantages of bits apply to all digital documents, whether they reside in a digital annotation system or not. These include (a) copies of digital documents are identical to originals and multiple copies can exist simultaneously (e.g., the original, one annotated by the student, a student-annotated version being viewed by the teacher, and one available for viewing by other teachers and parents); (b) additional information (e.g., annotation) can be added and modified as needed; (c) digital documents can be searched with sufficient speed to make looking up a word or phrase practical; and (d) digital files are transportable across networks of computers at the speed of light. Further, in digital annotation systems, hardware and software combine to keep all reading and annotation tools simultaneously ready for use. Readers use the same tool (i.e., a mouse, a stylus) to turn pages, browse, and add

annotations. This capacity, which Marshall (1997) called "ready-to-hand" (p. 5), streamlines the making of annotations to the point where it is easier to make annotations in some digital annotation systems than it is on paper. For example, digital annotation systems, because they allow readers to manipulate bits instead of atoms, permit readers to modify their annotations after they have been made (i.e., the color of highlighted text can be changed or the highlighting deleted).

This study used a qualitative case study design to gather data to answer the question "What are the outcomes of using digital annotation software among secondary school teachers?" Two high school teachers used RepliGo™ digital annotation software (Cerience Corporation, 2003) with their students over the course of one month during the fall of 2005. Teacher interviews, curricular materials, and student work were analyzed using Hughes' (2000) Replacement-Amplification-Transformation taxonomy to determine outcomes attributable to the use of RepliGo™.

Review of the Literature

The historical role of annotation in learning from text. Annotation at its minimum is the marking of an existing text. A *mark* can be any visible trace and a *text* could be letters and/or symbols, audio (Coates, 2005; Northwestern University, 2005), or video (Smith, Blankinship, & Lackner, 2000). Jackson (2001), in her seminal historical examination of annotations made in books, defined annotation as marks that are "the product of an interaction between text and reader" (p. 100). She described the gamut of annotation in books and included all possible marks regardless of purpose or intent (i.e., a signature on the inside front cover, underlining, written notes). Jackson's wide-ranging

analysis also included the role annotation plays in helping readers understand what they are reading.

The practice of annotating books as an aid to the construction of meaning has been in existence for centuries. Jackson (2001) found numerous recommendations to make annotations in educational treatises and concluded annotation is "a minor theme in educational theory" (p. 48). Teachers as far back as Erasmus in the early sixteenth century "advised pupils on methods of annotation and supervised and encouraged their efforts" (p. 46). Jackson cites Pryde, author of *Highways of Literature; or, What to Read and How to Read* in 1882, saying, "Pryde promotes a system of note-taking followed by written digests in the reader's own words" (p. 49). More recently, Frederick (1938) exhorted the student to "…develop a system of marking his books so that he will quickly see *what* his thoughts were as he reads" (p. 213). Adler (1942) and Adler and Van Doren (1972) called annotation *marking* and maintained that critical reading relied on it. Finally, a form of annotation called *functional underlining* was considered by Miller (1980) "as basic to the intelligent reading of literature as a test tube is to a chemistry experiment" (p. 577). Jackson found annotating directly in books had several advantages for learners over making notes in a notebook, among them (a) the writing of annotations requires less concentration and (b) an annotation allows the passage of text in the book to be referred to later as a check against the reader's interpretation. Despite the accepted usefulness of annotation to readers' efforts to understand what they read, there are few accounts of how useful students' annotations are to their teachers.

In contemporary post-secondary institutions, where annotation is a common practice among students, instructors seldom make use of students' annotations. Salvatori (1996) was an exception when she described how seeing a student's highlighted (equivalent to underlined) textbook altered her ideas about students' annotations:

> That class made it possible for me to turn a rather mechanical study habit—the highlighting of a text—into a strategy, one that can make visible the number and the intricacy of strands in a text's argument that a reader (or an interlocutor) pays attention to and that can show how the selection, connection, and weaving of those strands affects the structuring of the argument a reader constructs as/in response to a text (pp. 189-90).

This is a case where an instructor identified one of Jackson's (2001) advantages of annotation: the text itself served as a check against the student's interpretation. However, this was a chance, not intentional, encounter. If the student had not chosen to record her interpretation or if Salvatori's eye had not been caught by the vivid highlighting in the book open on the student's desk, then the alteration in instruction would not have occurred.

Literacy researchers' approaches to annotation research. Literacy researchers have studied two types of annotations: underlining and/or highlighting, and margin notes (Berger & Schlitz, 2000; Fowler & Barker, 1974; Hartley, Bartlett, & Branthwaite, 1980; Hynd, Simpson, & Chase, 1990; Lindner, Gordon, & Harris, 1996; Nist & Hogrebe, 1987; Nist & Simpson, 1988; Peterson, 1992). Participants in these eight studies were post-secondary students except those in the Hartley, Bartlett and Branthwaite study

(1980), which studied sixth grade students. All studies used experimental designs, and the dependent variable in each study was student performance on a recall measure.

The literature reviews in two of the studies (Hartley, Bartlett, & Branthwaite, 1980; Nist & Simpson, 1988) surveyed a larger number of studies than the eight studies examined for this review. Hartley, Bartlett, and Branthwaite (1980) reviewed forty-one studies on the effectiveness of underlining and found that "few studies, if any, provide clear-cut support for the effectiveness of underlining, and that little is known of how children benefit (or don't benefit) from underlining" (p. 218). Nist and Simpson (1988) reviewed 13 studies that either (a) compared the effectiveness of researcher-supplied underlining with student-generated underlining or (b) compared the effectiveness of underlining with other study strategies. They found "the only safe conclusion which can be drawn is that underlining does not to appear to be detrimental" (p. 251).

Underlining was the independent variable in three of the studies, with two (Hartley, Bartlett, & Branthwaite, 1980; Nist & Hogrebe, 1987) comparing researcher-provided underlining with student-generated underlining, while the third (Peterson, 1992) compared student-generated underlining against no underlining. Hartley et al. (1980) found that sixth graders who studied materials with researcher-supplied underlining performed significantly better on a cloze test. Nist and Hogrebe (1987) found no significant difference between the treatment and control groups, and Peterson (1992) found the control group outperformed the treatment group on an inferential recall measure.

7

When the independent variable was highlighting, Fowler and Barker (1974) reported no significant difference between groups of college students assigned to highlighting, underlining, or control groups. Lindner, Gordon, and Harris (1996) found no significant differences in students' performance on a multiple choice test in their first experiment, and in their second experiment, the control group outperformed the treatment group. Berger and Schlitz (2000) examined the effect of researcher-supplied underlining that was intentionally poor, comparing it to a group who received researcher-supplied highlighting that was high-quality, and a group that received no intervention. Berger and Schlitz found no significant difference between the two groups who received the intervention, and the control group outperformed both of the experimental groups.

Two studies (Hynd, Simpson, & Chase, 1990; Nist & Simpson, 1988) examined the role of margin notes in promoting recall as measured by performance on multiple-choice tests. In both studies, groups trained to make annotations in the form of margin notes outperformed either the use of self-selected study strategies or the use of a journal.

To summarize the literacy research on annotation, the primary annotation method investigated is underlining and/or highlighting, and the reason to annotate is to enhance performance on recall measures. The results indicate that, at best, the use of annotation to enhance recall produces no significant improvement. Some research does indicate recall improves when students are trained to annotate by making margin notes. However, none of the studies reviewed investigated annotation as a formative assessment. This represents a gap in how the field conceives of annotation, perhaps due in part to the fact that the use of students' annotations as formative assessments requires teachers to have access to their

students' annotations. This access is inherently impractical when annotations are made on paper materials. This study addressed that gap by placing a digital annotation system into the hands of K-12 teachers and reporting teachers' uses of their students' annotations as formative assessments.

Research into the development and uses of digital annotation systems. The research literature surveyed for this portion of the review focuses on the development and uses of digital annotation systems. Six studies (Marshall, 1997, 1998; Marshall, Price, Golovchinsky, & Schilit, 1999; Ovsiannikov, Arbib, & McNeil, 1999; Schilit, 1999; Waller, 2003) examined how adult readers annotated paper documents to inform the design of digital annotation systems. Two of these (Marshall, 1997; Ovsiannikov, Arbib, & McNeil, 1999) are reviewed here to explicate some of the ways learners use annotations.

Marshall (1997) analyzed college students' annotations by examining used textbooks selected from the stock of a college bookstore. Marshall did not have access to the students who annotated the books in her sample and thereby relied solely on the highlighted books when interpreting what she found. Marshall found that annotations functioned as place markers and aids to memory (i.e., short passages Marshall supposed were for later use in writing a paper or preparing for a test), records of interpretive activity (i.e., marks that indicated unfamiliar language, commentary, and evidence of misreading), and visible traces of the reader's attention (i.e., more annotations were present when content seemed more difficult to comprehend).

Ovsiannikov, Arbib, and McNeill (1999) surveyed postsecondary students and professionals in neuroscience and computer science to learn how and why they annotated paper documents. Their findings included the development of a taxonomy that described their participants' uses for annotations as they were being written and/or once they had been made. These were (a) *to remember*, where readers marked the main points of the document during reading and returned to these annotations at some later point in time; (b) *to think*, where readers made these annotations during reading when they thought of new ideas, questions, and opinions; (c) *to clarify*, where readers' annotations restated difficult concepts in their own words during reading; and (d) *to share*, where readers shared their annotations collaboratively for document editing and as the basis for conversations.

Marshall (1997) and Ovsiannikov et al. (1999) identified types of annotations learners made that could be useful as formative assessments. Those identified by Marshall (i.e., highlighting a word because it was difficult, highlighting a paragraph because it was identified as important) are metacognitive traces (Winne & Hadwin, 1998). A highlighted word or phrase could signal to the teacher a difficulty a student is having, such as not being able to understand a word or being unable to distinguish what is important in a paragraph. Types of annotation identified Ovsiannikov et al. (e.g., margin notes that document a student's questions and/or opinions) would be evidence of higher-order thinking skills.

Four studies (Lebow et al., 2004; Lick & Lebow, 2003a, 2003b; Nokelainen, Miikka, Kurhila, Floréen, & Tirri, 2005) examined uses of digital annotation systems in educational settings and indicated high levels of applicability to teaching and learning.

10

Nokelainen et al. (2005) found that students in their university-level, web-based courses thought a digital annotation system added value to the learning process, changed their study habits favorably, and would be useful in other courses.

Lick and Lebow (2003a; , 2003b) and Lebow et al. (2004) reported on a trial of their digital annotation system, Hylighter, which "records a trace of students [sic] reading activity" (Lick & Lebow, 2003a, p. 9). The trial occurred in a community college language arts course with an instructor who had taught the same course for five years. All students in the trial were high school graduates and had successfully completed the institution's introductory English course. During the trial, the instructor used the system to give her students formative assessments. The initial assessment "exposed serious reading comprehension problems including general confusion in following relatively simple instructions" (p. 9). This first experience with the system "initiated a process that has changed the instructor's approach to teaching" (Lick & Lebow, 2003a, p. 9). At the end of the trial the instructor reported student performance as determined by grades on assignments greatly exceeded expectations, and the course completion rate was 97% compared to the typical rate of 70%.

Methods

This section provides the methodological context for this study, including descriptions of the research design, participants, materials, procedures, and analysis.

Design. The study used a qualitative case study design (Yin, 1994). A qualitative case study design investigates phenomena within a real-life context and thus was deemed

appropriate for this examination of teachers' uses of digital annotation in their classrooms.

Participants. Two high school teachers from a large suburban school district near a large metropolitan area in the Southern United States volunteered to participate in the research. Richard and Julia (pseudonyms) were among a pool of secondary school teachers (N=123) in the district who attended one in a series of introductory workshops on the use of RepliGo™ digital annotation software held in late April, early May, and mid-August, 2005. The district was selected because all high schools and one middle school were *laptop schools*, where each teacher and student had been provided with a district-owned laptop computer for use at home and at school during the school year. Laptop schools are a subset of schools that offer *one-to-one computing*. One-to-one computing is defined as "each student and teacher has one Internet-connected wireless computing device for use both in the classroom and at home" (Greaves & Hayes, 2006).

Richard taught Spanish in primarily ninth and tenth grades at Adams High School (pseudonym). He became a teacher following a seventeen-year career in broadcast journalism and earning his Master's degree in Education. He was in his fourth year of teaching when he began the field trial.

Julia taught Reading at Madison High School (pseudonym) to ninth through twelfth-grade students who had failed to pass the required statewide reading assessment. Julia completed her Bachelor's degree and was nearly finished with a Master's degree in experimental psychology at the time of the field trial. She was in her third year of teaching and second year at Madison High School.

Materials. Materials included a participation incentive, the RepliGo™ digital annotation software, online support materials, a semi-structured telephone interview protocol, participant-provided curricular materials, and copies of student work. The incentive was a permanent license to the RepliGo™ digital annotation software provided by the publisher.

Digital annotation software. The digital annotation software used in this study was RepliGo™ version 2.0 (Cerience Corporation, 2003). RepliGo™ has three components: the standalone viewer software, a web-browser plug-in viewer, and a print driver that converts electronic files into RepliGo™ format so they can be used by the viewers. The RepliGo™ viewer software allows users to add highlighting and notes to electronic documents that have been prepared using the RepliGo™ print driver. Highlighting can be performed in one of eight colors, and annotations appear in a database displayed alongside the viewer window. RepliGo™ allows users to navigate rapidly between highlighted portions of the text by providing the option of viewing the text in a variety of magnifications.

Online support materials. A website was created to support Richard and Julia and their students. It included RepliGo™ software downloads, how-to movies and user guides, and RepliGo™ documents created to illustrate possible content-specific classroom uses of RepliGo™.

Semi-structured telephone interview protocol. Three weekly, semi-structured, individual telephone interviews were conducted with Richard and Julia during the month they used RepliGo™. Interviews allowed Richard and Julia to fully describe their uses of

RepliGo™ and any changes in their teaching practice they noticed as compared to how they taught prior to using RepliGo™.

Participant-provided curricular materials. Richard and Julia used the RepliGo™ print driver to convert self-selected curricular material from their existing curricula for their students to access and manipulate using RepliGo™. The materials included paper documents that were scanned and converted into RepliGo™ files, web pages and word processing files. Examples included Richard's story excerpts he used for vocabulary assessment and a short story Julia used with her students to learn about purpose when reading. Richard and Julia provided these files to the researcher, and they were reviewed during the telephone interviews. The researcher asked Richard and Julia about the files to more fully understand their uses of RepliGo™.

Examples of student work. Julia provided seven anonymous examples of students' work on the short story mentioned above. She introduced and shared these examples during a telephone interview to illustrate how she was using RepliGo™.

Procedures. The study consisted of two case studies of individual teachers, Richard and Julia, who used RepliGo™ digital annotation software in their classrooms. Data for the case studies included three individual, semi-structured telephone interviews and classroom artifacts (e.g., RepliGo™ files provided to students, examples of student work). Interviews were conducted in October, 2005, ranged in duration from ten minutes to forty-two minutes, and focused on the variety of Richard and Julia's uses of RepliGo™ across three broad themes of technology use developed by Hughes (2000): (a)

14

instructional method, (b) student learning processes, and (c) curriculum goals. Interviews were audio-recorded and transcribed.

Analysis. The case study data gathered were analyzed using Hughes' technology use taxonomy (Hughes, 2000) to answer the question, "What are the outcomes of using digital annotation software among secondary school teachers?" First, a list of instances of use of RepliGo™ was compiled from the transcripts of the interviews with Richard and Julia. An instance of use occurs when technology is used before, during, or after a lesson. Then each instance was coded using Hughes' (2000) themes of teaching practice (i.e., instructional methods, curriculum goals, student learning processes) and focus of use (i.e., teacher, student) as an a priori coding scheme. Next, Hughes' descriptions of categories of technology use (technology as replacement, technology as amplification, technology as transformation) were applied to each instance to determine the degree to which each theme "remained the same, was more efficient, or changed" (Hughes, 2000, p. 18). Hughes' categorization protocol was followed whenever an instance of use fell across multiple categories of use; the entire instance was placed in a category based on the most transformed use. An example was when Richard amplified his previous assessment practice because using RepliGo™ was faster and transformed his interactions with his students because he spent time with them one-on-one during the assessment where he did not do this at all before. This entire instance was categorized as a time when the technology transformed the instruction.

To evaluate the reliability of the coding, a competent second coder (a graduate student with experience coding data with RAT under the supervision of a faculty

member) coded 9 (16%) of the 58 statements coded by the researcher. Initial agreement was acceptable (78%) on both the coding of the theme and category of technology use. Subsequent conversation resolved all discrepancies.

Once the instances were classified according to taxonomy categories, two sets of rich descriptions of illustrative instances were created, one for Richard and one for Julia. These rich descriptions were used to illustrate, using the RAT taxonomy, the outcomes of using RepliGo™ generated by Richard and Julia.

Results

Richard's instances of use. Richard taught Spanish to both native and non-native Spanish speakers in primarily grades nine and ten at Adams High School (pseudonym). He reported 11 instances of using RepliGo™ during the four week trial. Nine instances occurred when his students used RepliGo™ to complete formative assessments of their vocabulary skills. While these instances were of the same use, they are considered distinct because Richard's lesson plan and materials varied each time. One instance occurred when he projected text for an in-class discussion, and the last instance occurred across several uses as he used RepliGo™ to prepare the formative assessment materials for his students.

Richard's use of RepliGo™ for formative assessments of his students' vocabulary skills was something he suggested as a possible use during his workshop the previous spring. He thought they could use the software to highlight words they did not know in passages from his existing curriculum materials. Richard had his students use RepliGo™ to complete formative assessments in nine unique instances (i.e., Richard used a different

16

lesson plan), and each instance was repeated two or three times (i.e., Richard taught the same lesson two or three times in a school day). During these instances Richard became aware of changes in his students' behavior and his interactions with them.

Richard mentioned numerous times that more of his students' participated in the formative assessment activity using RepliGo™ than they did when using other methods, and he was able to check their understanding more rapidly. He says:

> Before I use[d] to project the [Spanish] story… I'd say "OK, we're going to read and translate, we're not going to read it out loud in Spanish." But maybe five or six kids in the whole class out of twenty-five would be actively participating in reading and/or translating… and the others might… [be] watching, but…there's no way to know for sure if they're daydreaming… or what. But when you use it [RepliGo™] and you project it and you say, "OK now it's on your computer, please open [it] up… and scan through it and [highlight] what you don't know"…Then you know everybody is getting involved and you can check on that right away (Richard:1,6) .

Richard attributed students' increased participation to RepliGo™ being a technological tool, one that students felt comfortable with and enjoyed using, saying:

> They enjoy it, they enjoy the technology and they enjoy learning it and they're naturals at it so when I give them an assignment… I get pretty much 97-98% of the students [to] just eagerly and gladly and willingly do the assignment. I mean they just jump right into it. They appreciate doing it very much. And a lot of it is related to just the fact that it is a computer and it's colorful. (Richard:1,1-2)

17

These two comments, taken together, are an example of technology as amplification in that the use of RepliGo™ by the students in Richard's class increased their motivation and their participation in the assessment activity. This use falls within the theme of student learning.

Richard valued the way RepliGo™ enabled his students to complete formative assessments, saying "…that [using RepliGo™] was a new approach and again a very rapid, visual bit of feedback for the teacher about what the kids say they don't know. I mean how often do kids voluntarily say 'I don't know something'?" (Richard:1,5). While Richard includes language referring to an increase in the speed of the assessment, which would make this instance of use an amplification, this quotation is an example of a transformation, this time in the usage theme of instructional methods. RepliGo™, in Richard's mind, has produced a fundamental change in the willingness of his students to share with him what they need help with. Additionally, Richard now has access to such information that, heretofore, had been in the students' minds only.

Richard also credited RepliGo™ with changing his interaction with his students during the vocabulary assessments. He explained:

It [RepliGo™] changes it in that… I am interacting with them, whereas when they're taking the quiz in Blackboard™ [the district's electronic course management software], I'm just walking around the room just to make sure they're not cheating…But with this… I want them to tell me what they don't know… so I'm walking around looking at… a lot of different screens seeing what's being highlighted… I mean that kind of assessment produces a lot more

18

teaching… I'm saying you know, buscar means "look for," or I might gesture to them or something or other as a reminder…That wouldn't be happening with just a quiz. (Richard:1,6)

Near the end of the field trial Richard brought up this same theme of interaction with students again, saying:

So I think it's sort of put me moving around the room a little bit more, checking more directly on [students'] work, and... I think it's created that link and strengthened that particular activity because most of what I do is during class time, not afterwards, and I don't write lots of notes on papers… It's got to be during the class period that that's taking place. (Richard:3,7)

Richard used RepliGo™ to fundamentally alter his interactions with his students during formative assessments. Instead of moving about the room to prevent cheating, he gathered information from his students. Instead of issuing a quiz grade, he learned what his students did not know so quickly that he could assist them immediately. His interest in helping his students learn became more visible with this type of assessment made possible by RepliGo™. This transformation occurred early on and remained in place for the duration of the field trial.

Richard's second instance of use was to project a paragraph from a RepliGo™ file on his computer onto the projection screen at the front of his classroom. He then used RepliGo™'s highlighting tools to highlight parts of a paragraph in various colors as he was leading a whole class discussion on the meaning of the passage. He uses the term "overhead" to describe the hardware/software combination of computer, video projector,

and RepliGo™, saying "... I posted it [the paragraph] up on the overhead [and] highlighted with different colors key sections…" (Richard:4,1). This use of RepliGo™ was using technology as replacement. The usage theme applied to this instance was instructional methods.

Richard used RepliGo™ several times to prepare passages for his students' formative assessments. Mostly brief stories, usually one page long, these passages were from his existing paper curriculum materials. He scanned these and used optical character recognition software to import them into a word processor. Once he corrected any errors, he converted them into RepliGo™ documents and distributed them to students using Blackboard™. Sometimes Richard converted materials that were already in electronic format (i.e., word processor files, web pages). Applying the RAT taxonomy (Hughes, 2000) to this instance was not possible because during his interviews, Richard did not compare materials preparation using RepliGo™ with how he prepared his prior formative assessment materials, namely quizzes.

Julia's instances of use. Julia taught Reading at Madison High School to students who had not yet passed the statewide reading assessment, and she had approximately 140 students each day across six sections. Julia had three instances of using RepliGo™, all related to a single lesson she repeated with each section throughout the same day. She used RepliGo™ to prepare an assignment for the lesson, her students used it to complete the assignment, and Julia used RepliGo™ again to view her students' work while scoring it.

The goals of Julia's lesson were to demonstrate the value of having a purpose while reading and to familiarize her students with using RepliGo™. The lesson consisted of an assignment to read and highlight a single page of text using RepliGo™. The text described a house and its contents and included directions to highlight the text according to three purposes: what students found important, what a thief would find important, and what a potential home buyer would find important. Students were asked to highlight the text using a different color to identify each purpose. Julia did not create this assignment herself; she had previously obtained a paper version of it from a colleague. She had not used the assignment on paper before because she was not able to provide her students with the highlighter pens needed and thought that substituting other marks (i.e., underlining, circling) would be unsatisfactory.

Julia prepared the assignment by using RepliGo™'s file conversion software to convert a word processor file into RepliGo™ format, added the directions, and distributed the file to her students using Blackboard™. Students accessed the RepliGo™ file by logging into the class Blackboard™ site, where they worked with the file using RepliGo™'s browser plug-in software within their web browser. When they were finished, students used Blackboard™ to grant Julia access to their work. Julia then scored students' work by logging into Blackboard™ and viewing each student's RepliGo™ file inside a web-browser.

Out of the 140 students Julia taught that day, between 10 and 15 of them did not have their computers with them. Julia printed out the assignment and had these students complete it using a pen by underlining and circling the text. This unanticipated situation

21

provided a natural point of comparison for Julia between paper and RepliGo™, especially

when it came time for her to score the assignment.

Applying the RAT taxonomy to Julia's uses of RepliGo™, when Julia used

RepliGo™ to prepare the assignment and distributed it to her students via the

Blackboard™ course management system, she reported that the two software applications

streamlined the process "because I don't have to truck up to the copier and worry about

whether or not [the copier is] actually working and that is a pretty big chunk of

time…just to make copies" (Julia:1,7). RepliGo™ allowed her to make the worksheet

electronic rather than paper, saving the time needed to travel to the photocopier and

avoiding potential delays if the photocopier wasn't working. Blackboard™ let Julia

distribute the RepliGo™ file quickly to all of her students, thereby avoiding having to

pass out paper in class, which takes time. Because the result of using the technology was

streamlining the preparation process, this instance of use was categorized as technology

as amplification. Because it was Julia's preparation for teaching that was amplified, this

instance of use falls under Hughes' (2000) theme of instructional methods.

Julia credited RepliGo™ with making it possible for her to consider using the

assignment. Responding to a question on whether the availability of RepliGo™ affected

her decision to use the assignment, Julia responded,

> Well, yes, because I don't have enough highlighters that are three different colors
>
> for the kids. So it's very difficult to do… we can't go through and do the three
>
> perspectives, and if they just try to do it with pen or pencil and underline versus

circle, it just gets to be a mess and they can't really see the difference as clearly (Julia:1,4).

While the assignment was technically possible to do using highlighters and paper, for Julia it was possible to use technology to highlight text but not highlighter pens; they were a scarcer resource than laptops. Julia had actually considered using the assignment but had not done so until RepliGo™ became available. At the surface, this use of RepliGo™ as an assessment of students' ability to perceive purposes while reading seems to represent a change in Julia's instructional methods. However, upon closer examination, it is difficult to say what changed because we do not know how Julia taught this concept before RepliGo™ was available. If she had never taught this concept before and RepliGo™ played a central role in making it possible, then this instance would have been categorized as transformation. If she had taught the concept before and used a different assignment to assess it, then this use of RepliGo™ would have been categorized as a replacement or amplification. In reality we do not know the categorization of this instance simply because we do not know if, or how, Julia taught the concept in the past.

Julia's third instance of using RepliGo™ was to view her students' work after they had completed the assignment. She noticed that more students turned in the assignment, reporting that "I'm looking at about 90% actually turned it in versus on any... typical assignment maybe 75% will turn it in" (Julia:1,5). It was also while viewing the work that Julia noticed the differences in the work done by students using RepliGo™ and the 10-to-15 students who did not have their laptops with them that day. Those students did the assignment on paper using a pen, and Julia reported "…when I look at their

papers compared to the ones with [students' who used] RepliGo™, I'm just not so sure they [the paper-using students] got it as well" (Julia:1,5). So Julia noticed more students turned in this assignment and more students appeared to understand her goal of having a purpose while reading when they used RepliGo™. Because Julia reported an increase on her measures of student performance, but not a fundamental change, this instance of use of RepliGo™ is categorized as using technology as amplification, and because Julia perceived this instance when looking at student work, it is classified under the theme of student learning.

Summary of the analysis of the cases. Richard's use of RepliGo™ to conduct formative assessments of his students' vocabulary skills transformed his instructional methods and amplified his students' learning. Apart from his use of RepliGo™ to prepare the assessments, this was his first use, one he mentioned during his first exposure to the software. Using RepliGo™ to project a passage for whole class discussion came later, near the end of the field trial. This use was a replacement of a set of older technologies (i.e., overhead projector, transparencies, markers) with a set of computer-based technologies (i.e., computer, video projector, RepliGo™) with no change in Richard's instructional methods, his students' learning, or the curriculum. Julia's uses of RepliGo™ amplified her instructional method by making her preparation of learning materials more efficient, and amplified her students' learning because more of them completed an assignment when they used RepliGo™. Both of these instances occurred within her first and only week of using RepliGo™, and she did not report a use that could be categorized as a replacement. This sequence supports the finding in Hughes' (2000) study of

24

technology use among English teachers: Teachers' level of technology use does not follow a sequential order and is not based on experience with the technology. Neither Richard nor Julia began by using RepliGo™ as a replacement for existing tools and then later used it to amplify and finally transform their practice. Instead, they used RepliGo™ in pursuit of goals that mattered to them.

Discussion

In this section I discuss Julia's discontinuance of RepliGo™ after one day of use in her classroom. Implications of her discontinuance are connected to the literature on the effects of high stakes tests and teachers' adoption of innovative teaching methods. The section continues with a discussion of Richard and Julia's uses of RepliGo™ as contributions to the research on the formative assessment of reading comprehension.

Julia's discontinuance of RepliGo™. During her RepliGo™ workshop in late April of 2005, Julia was clear that using RepliGo™ would help her students prepare for the state-mandated reading assessment. However, three events in the timeline leading up to her participation decision and her subsequent use of RepliGo™ combined to reverse her perceptions of RepliGo™ and move her from adoption to rejection. The first event was the timing of her late April workshop just two weeks after her students had completed the state-mandated reading assessment for the 2004-2005 school year. In essence, the pressure of the state-mandated assessment was reduced, and she was free to consider using RepliGo™. Second, Julia did not have the chance to begin using RepliGo™ with her students until six weeks into the 2005-2006 school year. Third, the state-mandated reading assessment was moved earlier from mid-April to mid-January,

25

2006. Thus, instead of being able to implement RepliGo™ during a period *after* her students had completed the state-mandated reading assessment, Julia was implementing RepliGo™ less than three months *before* the state-mandated reading assessment.

The pressure on Julia to have her students pass the statewide mandated reading assessment was acute because her students had already failed the assessment once. She described this pressure during her final interview, stating she needed to use multiple-choice assessments instead of RepliGo™ because the statewide mandated reading assessment was primarily a multiple-choice assessment. Thus, the statewide reading assessment impacted Julia's adoption and discontinuance of RepliGo™ as a formative assessment. Julia's discontinuance of RepliGo™ supports Boardman and Woodruff's (2004) finding that "teaching in a 'high-stakes' assessment environment impacts the implementation, fidelity, and sustainability of new teaching methods" (p. 545). As stated earlier, it was Julia's initial perception of RepliGo™ as compatible with the goals of her class that was her principal reason for her participation. However, her perceptions of RepliGo™ changed when she compared the kind of formative assessment possible with RepliGo™ against her perceived need to use multiple-choice items as formative assessments. When Julia was making the comparison between using RepliGo™ and multiple-choice assessments, she may have been acting as Boardman and Woodruff described when they concluded "that some teachers may use 'high-stakes' assessments as their primary reference point in which to gauge the merit of innovative teaching practices" (p. 545). While Boardman and Woodruff (2004) examined the effects of high stakes assessments on the implementation of innovative teaching practices in general,

Frank, Zhao, and Borman (2004) included a school's emphasis on standardized testing as a factor affecting teachers' technology adoptions. Frank, Zhao, and Borman categorized standardized testing within their "job stress" category and noted "job stress can demand immediate resources, distract attention, and induce burnout, all of which may affect an individual's capacity and intent to implement innovations" (p. 157). Thus, Julia's discontinuance of using RepliGo™ is another example of the effects of high stakes assessments on teachers and, hence, their students. Thereby teachers implementing innovative assessments such as a digital annotation system must be given the opportunity to implement them in conditions where their practice is not so tightly constrained by accountability measures.

Richard and Julia's Uses of RepliGo™. In this section, I discuss how Richard and Julia's uses of the RepliGo™ digital annotation software contribute to the research on the formative assessment of reading comprehension. Three arguments for the expansion of research into digital annotation as a formative assessment are made, and four areas for future research are suggested.

Three fundamental arguments emerged from this study for expanding research into students' annotations as formative assessments of reading comprehension. First, annotation is an activity students do while reading, or as Snow (2002) stated, performing "operations to process the text at hand" (p. 15). Thus, using annotations as formative assessments brings the assessment closer to what Pearson (2005) calls the "'click' of comprehension" (slide 54), the moment when the student understands (or does not understand) the text. Using annotation reduces the teacher's reliance on artifacts such as

accounts of whether the student understood, what they understood, or quizzes on what they remembered. When Richard asked his students to highlight the words they did not know through annotation, he was substituting annotation practices for a quiz on what they understood. Richard stopped *asking* students to remember what they knew and began *seeing* what they did and did not know at the moment they encountered difficulty. Thus, digital annotation is an assessment that is delivered *while* students process the text, close to Pearson's "'click' of comprehension," not at some later point in time. Thereby digital annotation is a formative assessment closer to the moment of comprehension than other available assessments.

Second, students' annotations should be used as formative assessments because annotations constitute a metacognitive trace (Winne & Hadwin, 1998) left behind by the student as they read. For example, Julia asked her students to use RepliGo™ to demonstrate their ability to have a purpose while reading. As students read and annotated the passage describing a house and its contents, students monitored their thinking in order to notice when they shifted their purpose in reading among the three purposes the assignment demanded they use (i.e., the student's own purpose, the purpose of a thief, the purpose of a real estate agent). Thus, each time they annotated the text using a different highlight color for each purpose, they left behind a trace of the shift in their thinking. Winne and Hadwin called this activity "traces of study tactics" (p. 280) and recommended students' annotations as evidence for student learning. Thus, Julia was using her student's annotations as traces that documented their awareness of their learning. Thereby digital annotation is a visible trace of students' metacognition.

Third, annotation meets the standard for a useful assessment system for reading comprehension. Snow (2002) states that at a minimum, a reading comprehension assessment must demonstrate four strengths. It must: (a) be congruent with the processes involved in comprehension, (b) target operations involved in reading comprehension, (c) provide information useful for instructional decision making, and (d) provide transparent information (i.e., information useful to teachers with little technical training in assessment). Richard's use of RepliGo™ demonstrated these four strengths in that (a) students completed the assessment while reading, (b) students demonstrated they did or did not understand the vocabulary, (c) students provided information in their annotations that altered Richard's instructional method, and (d) Richard gathered this information after less than two hours of training. Thus, even though Richard's use of RepliGo™ as a formative assessment was simple, it was effective and met Snow's test for a useful assessment system.

The three arguments for expanding research into students' annotations as formative assessments pertain to *annotation.* However, it is the *digital* aspects of a digital annotation system that makes students' annotations practical formative assessments. In *Knowing What Students Know* (National Research Council, 2001), the authors include a chapter on the uses of information technologies to advance educational assessment. They conclude that information technologies remove constraints on assessment practice, extend "the knowledge and cognitive processes that can be assessed" (p. 288), and facilitate the embedding of assessment in instructional settings. RepliGo™, and digital annotation systems in general, remove two significant constraints: (a) the prohibition against

29

annotating in paper textbooks, and (b) the inherent impracticality associated with sharing annotations made on paper materials. When Richard observed words his students were highlighting, he was assessing a different cognitive process than when his students completed a multiple-choice vocabulary quiz, thereby extending the cognitive processes he was assessing. Before using RepliGo™, the cognitive process Richard was assessing when he used multiple-choice vocabulary quizzes was *supported recall* (i.e., students could study in advance and use the responses of the multiple-choice item as scaffolds to think through to the correct response). The cognitive process Richard was assessing when he was using RepliGo™ was *understanding a word in context*. Thus he extended the cognitive process from an inauthentic process (i.e., the multiple-choice vocabulary quiz) to an authentic process (i.e., reading a passage of text). Finally, Richard embedded the assessment in his instruction, using the information students provided him to help them *during* the lesson. Thus, Richard leveraged the digital aspects of RepliGo™ to advance his assessment practice, which then immediately transformed his instruction.

Richard's and Julia's uses of RepliGo™ constitute preliminary evidence that students' annotations are useful formative assessments of reading comprehension. While simple, the kinds of annotations Richard and Julia asked students to do were effective because the students' annotations provided Richard and Julia with information they could use to alter their teaching and help their students learn. However, the use of annotation for formative assessment is not limited to the ways Richard and Julia created during the month they used RepliGo™ (i.e., identifying unknown words, demonstrating having a purpose while reading). For example, Lick and Lebow (2003, 2003b) described their

Collaborative Annotation Model (CAM) and the Hylighter™ digital annotation system they created to enable teachers and students to collaboratively annotate an instructor-assigned reading. CAM is a five-step process whereby students are taught to annotate using techniques such as "question-generation and answer-elaboration…, student-generated elaborations of important points in the text, or…identify and comment on claims, supports, and logical fallacies" (p. 4). The techniques Lick and Lebow mentioned are more sophisticated than those generated by Richard and Julia and, when combined with Richard and Julia's, indicate a range of reading comprehension strategies that can be formatively assessed using students' annotations. However, the data provided by Richard, Julia, and the instructor in Lick and Lebow's study are merely a tentative beginning to the research needed to develop digital annotation as a viable tool for the formative assessment of reading comprehension. Thus, four areas are suggested for future research into the uses of digital annotation as formative assessments of reading comprehension. These are research into (a) the effectiveness of digital annotation as a formative assessment of reading comprehension, (b) the kinds of formative assessments that digital annotation systems can support, (c) how students learn to annotate and teachers learn to use their students' annotations as formative assessments, and (d) the conditions that must be in place so digital annotation can be used as a formative assessment.

References

Adler, M. (1942). How to mark a book. In R. S. Loomis (Ed.), *Modern English Readings* (4th ed., pp. 268-272): Farrar & Rinehart, Inc.

Adler, M., & van Doren, C. (1972). *How to read a book*. New York, NY: Touchstone.

Berger, I. P., & Schlitz, J. D. (2000). *On getting what you pay for: The detrimental effects of reusing highlighted books*. Paper presented at the American Psychological Association Annual Conference, Washington, D.C.

Boardman, A. G., & Woodruff, A. L. (2004). Teacher change and "high-stakes" assessment: what happens to professional development? *Teaching and Teacher Education, 20*(6), 545-557.

Cerience Corporation. (2003). RepliGo™. Fort Collins, CO: Cerience Corporation.

Coates, T. (2005). On the BBC Annotatable Audio project.   Retrieved October 31, 2005, from http://www.plasticbag.org/archives/2005/10/on_the_bbc_annotatable_audio_proje ct.shtml

Fowler, R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology, 59*(3), 358-364.

Frank, K., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education, 77*(2), 148-171.

Frederick, R. W. (1938). *How to study handbook*. New York, NY: Appleton-Century Company, Inc.

Greaves, T., & Hayes, J. (2006). *America's digital schools 2006: A five year forecast.* Paper presented at the Florida Educational Technology Conference, Orlando, FL.

Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference - sometimes. *Journal of Educational Research, 73*(4), 218-224.

Hughes, J. (2000). *Teaching English with technology: Exploring teacher learning and practice.* Unpublished Doctoral Dissertation, Michigan State University.

Hynd, C., Simpson, M. L., & Chase, N. D. (1990). Studying narrative text: The effects of annotating vs. journal writing on test performance. *Reading Research and Instruction, 29*(2), 44-54.

Jackson, H. J. (2001). *Marginalia: Readers writing in books*. New Haven, CN: Yale University Press.

Johnston, P. (2003). Assessment conversations. *Reading Teacher, 57*(1), 90-92.

Lebow, D. G., Lick, D. W., Marks, P., Peck, D., Wolfe, J. L., & Hartman, H. (2004). *Interactive annotation for teaching and learning*. Unpublished manuscript.

Lick, D. W., & Lebow, D. G. (2003a). Medieval scholarship meets 21st century technology part one: The collaborative annotation model. *Online Classroom, 2003*(June), 4-5.

Lick, D. W., & Lebow, D. G. (2003b). Medieval scholarship meets 21st century technology part two: The interactive annotation model. *Online Classroom, 2003*(July (revised December 21, 2003)), 4-5.

Lindner, R. W., Gordon, W. I., & Harris, B. R. (1996, April 8-12). *Highlighting text as a study strategy: Beyond attentional focusing.* Paper presented at the American Educational Research Association Annual Meeting, New York, NY.

Marshall, C. C. (1997). *Annotation: From paper books to the digital library* (Technical report). Palo Alto, CA: Xerox Corporation.

Marshall, C. C. (1998, March 24). *The future of annotation in a digital (paper) world.* Paper presented at the 35th Annual GSLIS Clinic: Successes and Failures of Digital Libraries, University of Illinois at Urbana-Champaign.

Marshall, C. C., Price, M. N., Golovchinsky, G., & Schilit, B. (1999). *Introducing a digital library reading appliance into a reading group.* Paper presented at the Fourth ACM International Conference on Digital Libraries, Berkeley, CA.

McKenna, M. C., & Stahl, S. A. (2003). *Assessment for Reading Instruction*. New York, NY: The Guildford Press.

Miller, J. W. (1980). Functional underlining: An essay in bibliography, criticism, and pedagogy. *College English, 41*(5), 575-578.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academies Press.

Negroponte, N. (1995). *Being digital*. New York, NY: Vintage Books.

Nist, S. L., & Hogrebe, M. C. (1987). The role of underlining and annotating in remembering textual information. *Reading Research and Instruction, 27*(1), 12-25.

Nist, S. L., & Simpson, M. L. (1988). The effectiveness and efficiency of training college students to annotate and underline text. In J. E. Readence & R. S. Baldwin (Eds.), *Dialogues in literacy research: Thirty-seventh yearbook of the national reading conference* (pp. 251-257): The National Reading Conference, Inc.

Nokelainen, P., Miikka, M., Kurhila, J., Floréen, P., & Tirri, H. (2005). A shared document-based annotation tool to support learner-centered collaborative learning. *British Journal of Educational Technology, 36*(5), 757-770.

Northwestern University. (2005). Spoken word project: ProjectPad. Retrieved November 12, 2005, from http://www.at.northwestern.edu/spoken/p04annotation.html

Ovsiannikov, I. A., Arbib, M. A., & McNeil, T. H. (1999). Annotation technology. *International Journal of Human-Computer Studies, 50*, 329-362.

Pearson, P. D. (2005). *Assessing reading comprehension and vocabulary: What does the research tell us? What should we do in our schools?* : Lecture presentation hosted by University of Minnesota Center for Reading Research, St. Paul, MN.

Peterson, S. E. (1992). The cognitive effect of underlining as a study technique. *Reading Research and Instruction, 32*(2), 49-56.

Salvatori, M. (1996). The "Argument of Reading" in teaching of composition. In B. Emmel, P. Resch & D. Tennery (Eds.), *Argument revisited; argument redefined* (pp. 181-192). Thousand Oaks, CA: Sage Publications, Inc.

Schilit, B. (1999). Why e-Read? Finding opportunities in the merger of paper and computers. *Future of Print Media Journal*(April), 1-9.

35

Smith, B. K., Blankinship, E., & Lackner, T. (2000). Annotation and education. *IEEE Multimedia*(April-June), 84-89.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND.

Snow, C. E., & Biancarosa, G. (2003). *Adolescent literacy and the achievement gap: What do we know and where do we go from here?* New York, NY: Carnegie Corporation of New York.

Tierney, R. J. (1998). Literacy assessment reform: Shifting beliefs, principled possibilities, and emerging practices (Distinguished educator series). *Reading Teacher, 51*(5), 374-390.

Waller, R. (2003). Functionality in digital annotation: Imitating and supporting real-world annotation [Electronic Version]. *Ariadne*. Retrieved April 29, 2005 from http://www.ariadne.ac.uk/issue35/waller/intro.htm.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in education theory and practice* (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage Publishing.