Computers & Education 82 (2015) 50–59

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Online assessment of strategic reading literacy skills

Laura Gil*, Tomas Martinez, Eduardo Vidal-Abarca

Estructura de Recerca Interdisciplinar de Lectura, University of Valencia, Spain

ARTICLE INFO

Article history:
Received 26 May 2014
Received in revised form
28 October 2014
Accepted 31 October 2014
Available online 7 November 2014

Keywords:
Reading assessment
Reading literacy
Strategies
Online reading

ABSTRACT

This study investigates the possibility of assessing strategic reading literacy skills with computers. The critical value of this assessment is the recording of online indices of the reader's behavior that can be interpreted in terms of strategies. The study uses materials of a standardized paper-and-pencil reading literacy test called *CompLEC* (Llorens et al., 2011) and a technology called *Read&Answer* (Vidal-Abarca et al., 2011) that presents texts and questions with a masking procedure that allows the recording of reading time and readers' actions to develop a computer-based version called *e-CompLEC*. We found that reliability and validity of the two versions are largely equivalent, and that e-CompLEC provides self-regulation and reading behavior indices predictive of performance. The study also shows how self-regulation is an important component of reading literacy processes.

1. Introduction

Computers open new possibilities for the assessment of reading literacy skills. They not only minimize the cost and the lag time in scoring by providing immediate feedback of results and reduce costs of test production (Bennett, 2002, 2003; Olson, 2003; Paek, 2005), but they also allow recording of the reader's behavior, which can be interpreted in terms of strategic reading. For instance, computers can record how much time the reader reads a document or the reader's behavior when rereading the document to search information for answering questions (Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Vidal-Abarca, Mañá, & Gil, 2010). These strategic decisions have an impact on performance, and they can be used for assessing students' reading literacy skills (Rouet, 2006). A major goal of this paper is to present a study performed with a prototype of a computer-based reading literacy test—called e-CompLEC—that combines the potential of computers with a software called *Read&Answer* (Vidal-Abarca et al., 2010), to provide this sort of assessment for high school students. As far as we know, no other reading literacy test of this type exists.

e-CompLEC is a prototype of a computer-based version of a standardized paper-and-pencil test, called CompLEC, designed to assess students' reading literacy (Llorens et al., 2011). They both include the same materials (i.e., texts and questions) but differ in the medium through which the tests are delivered and, consequently, the possibilities for recording readers' strategic behavior online while they are reading a text and answering questions. Thus, an additional goal of this paper is to test the equivalence of the electronic prototype and the paper-and-pencil versions of the test regarding their reliability and validity.

In this introduction, we discuss basic issues about three topics; i.e., the assessment of reading literacy skills, the use of computers for the assessment of reading skills and the possibilities of *Read&Answer* technology to evaluate students' reading literacy skills online.

1.1. The assessment of reading literacy skills

High school students are expected to use a vast array of written documents to perform either academic tasks (e.g., learning factual or conceptual information) or non-academic tasks (e.g., searching for specific information about hobbies or events of personal interest) (OECD, 2010). The use of documents to perform specific tasks involves reading literacy skills that go beyond classical reading comprehension skills, such as making inferences or macrostructure formation (e.g., Graesser, Singer, & Trabasso, 1994; Kintsch, 1998) as it demands that the reader

* Corresponding author. Department of Developmental and Educational Psychology, University of Valencia, Avenida Blasco Ibáñez 21, 46010 Valencia, Spain. Tel.: +34 963 98 38 48; fax: +34 963 86 46 71.
E-mail addresses: laura.gil@uv.es (L. Gil), tomas.martinez@uv.es (T. Martinez), vidala@uv.es (E. Vidal-Abarca).

http://dx.doi.org/10.1016/j.compedu.2014.10.026
0360-1315/

self-regulate her reading strategies to fulfill the requirements of the task (see Rouet, 2006 for a comprehensive discussion of this issue). Therefore, for example, the reader has to make decisions about what and how to read the text (i.e., skimming the text or performing a detailed and deep reading) when she needs to reread part of the document or when to stop reading because her goal has been achieved. Capturing these reading literacy skills is the goal of international assessment programs promoted by OECD for adolescents (PISA, Programme for International Students Assessment, OECD, 2009) and adults (PIAAC, Programme for International Adult Assessment Competencies, OECD, 2012). Besides, the role of tasks in combination with the two classical elements of reading (i.e., the reader and the text) is a central focus of new theoretical approaches about reading comprehension (Snow & RAND Reading Study Group, 2002). Furthermore, a recent review of seven reading comprehension computational models by McNamara and Magliano (2009) points out that none of them have considered the role of tasks, which is a clear limitation given that reading occurs in a social context in which tasks play a crucial role.

A key feature of this approach, which we call task-oriented reading or functional reading (Vidal-Abarca et al., 2010), is that the reader has to perform a specific reading task (e.g., read an article of a newspaper about the Asian tsunami in 2004), with a specific goal (e.g., answer comprehension questions about the causes of tsunamis). This task has two important consequences. First, some of the information within the text is relevant for the task (e.g., explanations about the tsunami that devastated parts of Southeast Asia), whereas other information may not be task-relevant (e.g., survivors' testimonies); consequently, the reader has to distinguish between relevant and less-relevant or irrelevant information to achieve her goal (Anmarkrud, McCrudden, Bråten, & Strømsø, 2013; McCrudden & Schraw, 2007). In other words, the text as a whole is not read in a homogenous fashion; furthermore, readers may not read the text entirely as instead there are texts elements more and less relevant for the task that affect readers' visual attention to different parts of the text (Kaakinen & Hyönä, 2011). Second, the reader has to make decisions about: (a) how to read the text initially; that is, either fully and carefully before moving to questions; (b) when referring back to the document to perform the task (i.e., answer a question about the causes of tsunamis); and (c) when to stop reading because the task has already been performed. Therefore, task-oriented reading demands specific self-regulatory processes beyond classical comprehension processes that include readers' metacognitive competencies; i.e., the ability to use a variety of appropriate strategies when processing texts and the ability to monitor and adjust their reading activity for a particular goal. Theories and models of comprehension have provided insights on how readers process a text, but they have not explicitly addressed the self-regulatory process of going back and forth from the text to the task, and vice versa, to perform a typical reading literacy task. There is evidence that, when readers have an available text to answer questions, the reader's strategies and processes are partly different from those which readers activate when the text is not available (Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Schroeder, 2011). The key point is the back-and-forth reading from the text and the task, and vice versa, which demands self-regulation and decisions on the part of the reader.

In PISA tests (OECD, 2010) as well as in CompLEC (Llorens et al., 2011), students are asked to use a text to answer a number of questions. So, students are free to read the document entirely and carefully and then read the questions, or just to skip the text and go directly to the questions, or even skim the text. Depending on these strategic decisions, the reader may decide to reread the text to answer any questions and then stop reading because she considers that the question has been responded to (Vidal-Abarca et al., 2010). These kinds of strategic decisions may have an impact on performance and can be recorded online, as we will show later. Therefore, reading literacy situations such as PISA tests, or task-oriented reading, coincide with theoretical approaches (McNamara & Magliano, 2009; Rouet, 2006; Snow & RAND Reading Study Group, 2002) pointing out some important limitations of current models and theories of reading.

1.2. Computer-based reading assessment

Computer-based models of reading assessment are rapidly expanding in several educational tests. Proof of this is that most of the major testing programs, such TOEFL, GRE and K-12 are now delivered through computer-based testing. Furthermore, PISA 2018 will only be delivered as a computer-based test. There are more than a few reasons for developing and implementing computer-based reading assessments (Bennett, 2002, 2003; Olson, 2003; Paek, 2005): they minimize the cost and lag time in scoring by providing immediate feedback of results; they reduce costs of test production; and, what is most interesting for us, they allow for analysis of students' reading processes that cannot be studied from paper tests alone. Additionally, recent studies indicate that high school students are familiar with computers and feel comfortable using them (Higgins, Russell, & Hoffmann, 2005; Kim & Kim, 2013), which also supports the implementation of computer-based assessment.

As more reading tests are delivered in a computer-based manner, the issue of how testing modes may affect students' scores and the tests' psychometric properties becomes increasingly important. According to the American Psychological Association's Guidelines for Computer-Based Tests and Interpretations (1986), when interpreting scores from the computerized versions of paper-and-pencil conventional tests, studies of the equivalence of the scores from the two different versions are needed to allow the continued use of normative and validity information developed from the paper-and-pencil version. To date, a substantial body of research has examined the differential effects due to the use of computer-based tests (CBT) instead of paper-and-pencil tests (PPT) (i.e., test-mode effects), often with mixed results. Some studies have found lower scores on CBT compared with PPT (Kim & Kim, 2013; Mazzeo, Druesne, Raffield, Checketts, & Muelstein, 1991; Russell, 1999), while others have found higher scores on CBT compared with PPT (Clariana & Wallace, 2002; Pomplun, Frey, & Becker, 2002), whereas other studies have reported no difference between PPT and CBT (Bodmann & Robinson, 2004; Holzinger et al., 2011; Mason, Patry, & Bernstein, 2001). Researchers have discussed various hypotheses to explain such mixed results. For instance, Noyes, Garland, and Robbins (2004) found that more effort is needed to complete CBT, suggesting that workload may be set as a test mode effect that needs to be taken into account when administering computer-based assessments, whereas Mason, Patry, and Bernstein (2001) found that differences in flexibility between CBT and PPT (e.g., some computer interfaces do not allow the student to skip, review, and/or change answers) are primarily responsible for test mode effects.

Related to flexibility, and focusing specifically on existing test-mode studies comparing PPT and CBT versions of reading literacy tests that include extensive reading passages, which is a feature of CompLEC and e-CompLEC, most results show lower performance on CBT than on PPT (Murphy, Long, Holleran, & Esterly, 2003; O'Malley et al., 2005). According to Paek (2005), these differences may be due to issues related to scrolling and the strategies that students use to organize information (e.g., underlining key phrases). However, these differences may disappear as students' familiarity with reading on computers increases and as computer interfaces include more tools to enhance student's reading comprehension.