Journal of Educational Psychology

The Role of the Scientific Discovery Narrative in Middle School Science Education: An Experimental Study

Diana J. Arya and Andrew Maul

Online First Publication, April 30, 2012. doi: 10.1037/a0028108

Journal of Educational Psychology

0022-0663/12/$12.00 DOI: 10.1037/a0028108

The Role of the Scientific Discovery Narrative in Middle School Science Education: An Experimental Study

Diana J. Arya and Andrew Maul
University of Oslo

In an experimental study ($N = 209$), the authors compared the effects of exposure to typical middle-school written science content when presented in the context of the scientific discovery narrative and when presented in a more traditional nonnarrative format on 7th and 8th grade students in the United States. The development of texts was controlled so as to isolate the presence of the discovery narrative structure as the independent variable; outcome measures were developed according to the BEAR Assessment framework to be sensitive to a range of levels of understanding of presented information and to focus only on the conceptual material presented in the texts. Students exposed to the scientific discovery narrative performed significantly better on both immediate and delayed outcome measures. These findings are discussed in the context of a larger argument for the inclusion of the scientific discovery narrative in science instruction.

Keywords: literacy, genre, science education, narrative

Much of the research in the field of science literacy has focused on the cognitive importance of having school science texts students find interesting and relevant. Do attempts to increase the interest value of texts serve to promote or distract from learning of conceptual information—or does it not matter? One possible method of increasing students' interest in educational texts is through the use of narrative; however, evidence concerning how it may be best used in the communication of scientific information is equivocal. In this study, we focused on the presentation of scientific information though a particular form of narrative text, termed the scientific discovery narrative (SDN). We position the SDN within the theoretical view of narrative as a universal, culturally embedded concept that people use, intentionally or not, to tell real or fantastical accounts of experience for a variety of purposes (Bruner, 1990; Mishler, 1995; Ricoeur, 1980; White, 1980). In this sense, the SDN is the experiential account of a scientist's discovery of new knowledge about natural systems. This study investigates the educational effects of texts that present scientific information encased in a narrative account of the scientific discovery process (the SDN), compared to more traditionally-styled texts that present the same conceptual information in straightforward expository format.

Science educators have long studied and argued for the usefulness of historical narratives in science education (e.g., Conant, 1957; Klopfer & Cooley, 1963; Metz, Klassen, McMillan, Clough, & Olson, 2007; Solomon, Duveen, Scot, & McCarthy, 1992), yet all such investigations have been in the form of case study research. Science educators such as Avraamidou and Osborne (2009) make a case for the usefulness of "mixed texts," which are defined as narrative texts with embedded descriptive or explanatory texts about scientific concepts or processes, for "making science meaningful, relevant, and accessible to the public" (p. 1683). These researchers argue that narrative is largely neglected in science learning and instructional practice, and they call for more rigorous research studies designed to investigate particular effects of narrative in science learning. To our knowledge, there have been no randomized experimental designs comparing educationally relevant outcomes between students who have and have not been exposed to narratives.

Some findings from literacy research suggest an alternative perspective on the role of narrative texts in science education. Hidi, Baird, and Hildyard (1982), for example, found that the mixing of genres as suggested by Avraamidou and Osborne (2009) can cause confusion when readers attempt to use genre-specific strategies for understanding key information within a text. Further, Hidi et al. (1982) found in a sample of school-related texts that while mixed texts were interesting (and thus memorable) to readers, the conceptual information contained in embedded narratives had little or nothing to do with the story. Similarly, Wade and Adams (1990) considered the narrative aspects of these mixed texts to be superfluous and seductive, serving only to distract readers from concept-specific details. Jetton (1994) also found that in immediate and delayed tests of recall of text material, readers demonstrated a stronger ability to recall story idea units (i.e., familiar, personal events that do not contain factual information) than informational idea units (key conceptual or factual information).

Similarly, other scholarship on the use of stories in science literacy research has focused on the potential for young readers to develop misconceptions from fictionalized, often anthropomorphized accounts of scientific concepts and processes that contain entertaining yet conceptually irrelevant information (e.g., Jetton, 1994; Madrazo, 1997; Mayer, 1995; Rice, 2002). Empirical re-

Diana J. Arya, National Centre for Science Education, University of Oslo, Oslo, Norway; Andrew Maul, Unit for Quantitative Analysis in Education, University of Oslo.

Correspondence concerning this article should be addressed to Diana J. Arya, 1503 Martin Luther King Jr. Way, Berkeley, CA 94709. E-mail: djarya@cal.berkeley.edu

search on the presence of seductive details in expository texts, for example, suggests that readers often retain less conceptual information when irrelevant yet emotionally entertaining details are present (Garner, Brown, Sanders, & Menke, 1992; Garner, Gillham, & White, 1989). However, narrative as a concept (and the SDN in particular) does not necessarily involve seductive, concept-irrelevant details.

The experimental texts designed for this study did not include the seductive details that are often present in biographical accounts of scientists (Allchin, 2003; Westerlund & Fairbanks, 2004). Instead, the reader vicariously experiences the scientists' processes of discovery, generally beginning with an interest in knowing more about the natural world and in building on the existing knowledge of other prominent figures (mentors, colleagues) and including accounts of unforeseen events (accidental experiments or chance encounters) and reactions to them, as well as the methods (planned experiments or systematic observations) that eventually led to discoveries.

SDNs, in this view, are science narrative texts that make the author visible. Although Marie Curie's (1904) *Radioactive Substances* is a highly technical volume written for the scientific audience, she does not stray from the overall structure of telling her story as a personal narrative of discovery. Findings on the effects of a *visible author* (a term that, in general, denotes academic texts that are written in the first person with a conversational style; e.g., Nolen, 1995; Paxton, 1997, 2002) suggest that readers are more engaged, reflect more, and have a greater awareness of the intended audience when the author has made herself visible. Traditional school science texts, on the other hand, are often characterized as formal, authorless descriptions of concepts and natural processes, often involving challenging academic language (Glynn & Muth, 1994; Koch & Eckstein, 1995; Norris & Phillips, 2003; Snow, 2010). Some have noted that these traditional texts are often devoid of a sense of the feelings of importance and intrigue that originally inspired the discovery of the scientific knowledge being described (Bowker, 2005; Latour & Strum, 1986; Leite, 2002; Lemke, 1990; Niaz & Rodriguez, 2000).

The SDNs used in the present study are not written in the first person, but do highlight the original authors (Marie Curie and Galileo Galilei, respectively) as the protagonists of the discovery account.

Another theoretical perspective that is relevant to many kinds of narratives is the *personalization principle*, which states that students learn more deeply when a text invokes a heightened social presence through (in particular) informal, conversational writing styles (Mayer, 2009). Making the author visible certainly may be a way of personalizing texts; in addition, the use of first person *I* constructions and addressing the reader directly using second person *you* constructions can foster more intimate connections between readers and texts, which have been found to have positive cognitive effects in several experimental studies (McLaren, Lim, Gagnon, Yaron, & Koedinger, 2006; Moreno & Mayer, 1999, 2002; Reeves & Nass, 1996). Greater personalization may be one pathway by which narratives in general and SDNs in particular can promote deeper learning; however, personalization is not a defining feature of narrative, as nonnarrative texts can also be personalized. As we hypothesize that the SDN will have a positive impact on learning through the greater student–text engagement encouraged by the account of the process of discovery, and not just through greater personalization, the degree of personalization was maximized in both the narrative and nonnarrative texts used in the present study.

An individual's interaction with a text is, of course, a product of both features of the text and features of the individual. Cultural differences and ethnic identity have well-documented effects on text comprehension; for example, linguists and discourse analysts have found that cultural assumptions and prior knowledge have significant influences on text interpretation (e.g., Murata, 2007; van Dijk, 2001; Widdowson, 2004). Psychologists have found that cultural beliefs and practices have an effect on the way that individuals construct knowledge (e.g., Greenfield, 1997; Lipka, 1991) and that self-conceptions of academic ability reflect cultural stereotypes that can further debilitate (or increase) academic performance (Steele & Aronson, 1995).

School culture and climate have also been found to affect students' academic performance (MacNeil, Prater, & Busch, 2009; Stewart, 2008; Wang, Haertel, & Walberg, 1997; Welsh, 2000). The quality of communication and morale among students, teachers, and administrators within a school, for example, can make a significant impact on the cognitive abilities demonstrated by students (MacNeil et al., 2009). Results from a longitudinal study by Lassen (2007) revealed that the number of reported suspensions in two schools predicted reading and math performance (i.e., standardized assessments). Similarly, Simmons and Blyth (1987) found that increased number of school suspensions were highly associated with overall low grade-level performance (i.e., grade point average).

Prior knowledge is another well-known predictor of text comprehension. Specifically, the more a reader knows about the concepts and words in a science text, the easier this text will be to comprehend (e.g., Chi, Feltovich, & Glaser, 1981; Kendeou & ven den Broeck, 2005, 2007; Recht & Leslie, 1988). Based on evidence of grade-related differences in exposure to the conceptual information targeted in this study, we also seek to investigate grade-specific effects of the SDN.

Thus, our primary research question is: What effect does the presentation of scientific information encased in the SDN (rather than through traditional expository text) have on middle school readers' ability to understand and remember this information? As a secondary question: Is there any evidence that such effects differ based on students' backgrounds, including such factors as their level of prior knowledge and school culture and climate?

Method

Participants

The sample consisted of 209 participants, of whom 107 were female and 102 were male; 106 (51%) were Caucasian, 73 (35%) were Latino, 20 (10%) were African American, four (2%) were Asian American, and six (3%) were other, mixed, or unreported. According to state standardized tests, 82% (171) of the students were performing at or above proficiency in English language arts. Students were in the seventh and eighth grades (ages 12–14) and were drawn from nine classrooms in two schools from separate school districts in northern California. The two schools were recruited through an online advertisement. Initially, 248 students returned signed letters of personal and parental consent; of these, 209 were present during the first session of the study and were thus included in the analysis. Seventeen students were absent from the second session, leaving a total of 192 participants present during both sessions. Students not present during the second session are excluded only from the analyses involving the second testing occasion.

Four of the nine classrooms (64 students) came from Oaks Middle School (pseudonym), a high-poverty (95% free and reduced-price lunch) urban school with a predominately Latino and African American student population and a reported school suspension rate of 26 per 100 students (above the reported district average). The remaining five classrooms (145 students) come from Ocean Middle School (pseudonym), a low-poverty (37% free and reduced-price lunch) suburban school with a predominately Caucasian student population and a reported school suspension rate of nine per 100 students (below the reported district average).

The seventh grade teachers from both schools reported that their students had received very little or no academic exposure to the concepts presented in the present study's texts. The eighth grade teachers noted that their students had received some instruction on the periodic table, chemical reactions, the structural aspects of elements, and Marie Curie's discovery of radium. According to the eighth grade curriculum overview, all students in this grade at Ocean Middle School are expected to have a solid understanding of most of the key conceptual information presented in the experimental texts about radioactivity.

Classroom observations (conducted prior to the administration of the study) corroborated teachers' reports of students' prior exposure to targeted conceptual information. Immediately prior to the study, the seventh grade participants at Ocean Middle were preparing and presenting semester-long projects related to drug abuse, while their peers at Oaks Middle were finishing a unit about evolution and beginning a new unit about reproduction. By contrast, the eighth grade participants at Ocean Middle were completing a unit about chemical reactions and conducting a final experiment that tested students' understanding of the effects of heat on elemental bonding. Thus, it could be inferred that the eighth grade students had recently been exposed to greater levels of targeted conceptual information, particularly for the *Radioactivity* text (one of the two texts developed in this study).

Materials

Texts. Four texts were created for this study: two versions for each of two topics (titled *Radioactivity* and *Seeing at a Distance*). For each topic, an SDN version and a more traditional expository version were created. The *Radioactivity* texts cover the concepts of elements, radioactivity, and chemical reactions, as well as some applications of these concepts to technology. The *Seeing at a Distance* texts cover the Galilean telescope and its role in the acquisition of knowledge about the Earth's moon and the moons of Jupiter.

The SDN and expository versions of each topic were created so as to be maximally similar in all respects, save for the presence of the discovery narrative as the mode of presentation in the former texts. Both text versions of each topic contain the same informational content, drawn from the California state standards (California State Board of Education, 1998; National Committee on Science Education Standards and Assessment & National Research Council, 1996) for seventh and eighth grade science education.

Length of texts. The SDN versions of the experimental texts were longer than the non-SDN versions, by 255 words for *Radioactivity* and 122 words for *Seeing at a Distance*. The additional text for the SDN versions was necessary to ensure that the same amount of conceptual information was present in both text versions. Conceptual mapping of all texts through an external review process helped ensure that the additional text was purely narrative based and did not contain concept-specific information. For example, the SDN version of *Radioactivity* includes Marie Curie's excitement about the mysterious glowing rock shown by her colleague, Henry Becquerel, while the non-SDN makes no mention of Curie's emotive state.

To avoid a straw-man comparison of an interesting and uninteresting text, both narrative and expository text versions were written with the intent of maximizing the engagement and interest of the reader. For example, both text versions of each topic begin identically, addressing the reader in a conversational manner, using first-person and second-person pronouns, with questions to consider and with a small amount of background information. For example, both versions of the *Radioactivity* texts begin thus:

> What are the basic elements that make up the earth? If you could break down a rock into its smallest parts, what would you find? Gold? Copper? Elements are individual pieces of matter that combine with each other and make up everything we see around us. Most of what we see and use in the world, such as air and water, is not made of one single element. For example, sodium and chlorine are two different elements that make up the salt that we use for cooking.

The difference between the two text versions is found in the mode of presentation of the science content. Whereas the expository versions present information in a straightforward manner consistent with standard seventh grade science texts, the SDN versions describe the activities of the scholars (Galileo and Marie and Pierre Currie, respectively) and the manner in which they came to understand this content. As an example, two passages from the expository and narrative versions of the *Seeing at a Distance* texts follow. From the expository version:

> And with this simple, powerful tool [Galilean telescope], we can see many details when we use it to look up into the night sky. The moon may look like a smooth ball of light covered with dark spots, but on a closer look through this telescope, we can see deep valleys and great mountain ranges. Through the telescope, we can now see all the different marks on the moon's surface.

Whereas in the narrative version, the passage emphasizes Galileo's process of discovery:

> When Galileo looked through his new telescope, he could see the surface of the moon, and so he began his first close look into space. He slept during the day in order to work and see the moon at night. Many people thought that the moon was a smooth ball with a light of its own. Now that Galileo had a closer look through his telescope, he realized that the moon's surface had mountains and valleys.

Both passages communicate information about the surface of the moon and how the Galilean telescope is a tool for investigating the night sky. Similarly, the two following passages from the expository and narrative versions of the *Radioactivity* texts describe the robustness of radioactive elements. From the expository version:

> The radioactive elements in pitchblende rock are powerful enough to withstand most acids, even at the highest temperatures. However, nonradioactive elements like iron, carbon, and uranium will react with different types of acids and either turn into a gas and dissolve away or turn into whole solid clumps. On the other hand, radioactive elements will remain unchanged.

The narrative version includes the process that led Marie and Pierre to understand the same information described above:

> Like true chemists, Marie and Pierre burned the pitchblende at different temperatures and added different kinds of acid to see what would happen. If they burned the rock too quickly or added too much acid, all of the pitchblende would be gone or destroyed, and they would have to start all over again. Marie learned from their experiments that the radioactivity stayed even after many other elements were burned away. She also learned that the radioactivity would not react with most acids. Even when she and Pierre added different acids to the pitchblende, the radioactivity was unchanged.
>
> The nonradioactive elements, like iron and carbon, would react with acid and either turn into a gas and dissolve away or turn into whole solid clumps, which could be removed from the pitchblende.

It should be noted that while uranium is mentioned in the expository excerpt, this element is not described until later on in the SDN version so as to be more consistent with the manner in which Marie and Pierre's discovery unfolded; thus, it is not present in the above excerpt.

Both text versions for each topic contained identical illustrations (11 total in the *Seeing at a Distance* texts and six total in the *Radioactivity* texts). All texts were organized and published in the form of small booklets.

Quality control. For purposes of text review and quality assurance, a 10-person external panel was formed, consisting of a chemist, two applied physicists, a professor of children's literature, a professor of psycholinguistics, a middle school librarian, and four middle school science teachers. The external panel members evaluated the texts for content accuracy, clarity and coherence of presentation, readability and text complexity, and for consistency across versions in these features and density of presented information.

Additionally, texts were piloted with 15 middle school students (five each from sixth, seventh, and eighth grade; eight males and seven females; four Caucasian, six African American, four Asian American, and one biracial); 11 of them were from standard classrooms, three were from special education classes, and one was from a gifted-and-talented classroom. Each student took approximately 20–30 min to read the SDN version of one topic and the expository version of the other topic (varied between interviews), with two students receiving all four texts. Comments given by students guided revisions of all texts in terms of coherence and accessibility, engagement, appropriate cognitive load, and overall presentation.

Text drafts were submitted for analysis by the software program Coh-Metrix (McNamara, Louwerse, Cai, & Graesser, 2005) for general characteristics of text coherence, including readability, word and text organization, syntactic complexity, referential cohesion, and clarity within four "situation model" dimensions. The Coh-Metrix indices indicated that the narrative and nonnarrative versions of both topics were extremely similar in all measures of coherence and readability. Three indices in particular (shown in Table 1) are worth noting: total number of words, the Flesch-Kincaid readability level, and latent semantic analyses (LSA).

Readability indices indicate that while the *Radioactivity* texts may be more challenging than the *Seeing at a Distance* texts (presumably due at least in part to the low general frequency of words like *radioactivity*, *electrometer*, and *radium* compared with words like *lens*, *convex*, and *rotation*). The LSA values (which refer to the "conceptual similarity" of words, sentences, and paragraphs within a text) of the narrative versions were consistently, albeit only slightly, lower than their nonnarrative counterparts. These results seem to align with the stance that narrative portions in mixed texts are unrelated or irrelevant to the expository information, thus providing some evidence for the notion that narratives in mixed texts introduce irrelevant and potentially seductive details.

Both the aforementioned external panel and the 15 students were asked for feedback regarding the texts. Of the four teachers involved in the editing of the texts, only one commented that the SDN versions would elicit greater interest and engagement on the part of the students; the remaining three commented that all texts had an equal chance of promoting such engagement. Of the sample (N = 209) of students described below, 105 received the SDN version of *Radioactivity* and the expository version of *Seeing at a Distance*, and 104 received the expository version of *Radioactivity* and the SDN version of *Seeing at a Distance*. Of the former group, 70 (67%) indicated that they preferred the SDN (*Radioactivity*) version; of the latter group, 29 (47%) indicated a preference for the SDN (*Seeing at a Distance*) version. These results show that the topic of *Radioactivity* was slightly more popular overall than *Seeing at a Distance* ($\chi^2_{(1)} = 8.27$, $p < .01$), which makes sense given some of the human mortality elements included in the discussion of radioactivity, and that the SDNs were only slightly more popular overall than the expository texts (with 119/209 students preferring SDNs). Thus, it seems reasonable to conclude

Table 1
Coh-Metrix Indices

Title	Word count	READFKGL	LSAassa	LSApssa	LSAppa
Radioactivity (non-SDN)	958	10.081	0.484	0.39	0.653
Radioactivity (SDN)	1,213	10.103	0.427	0.379	0.649
Seeing at a Distance (non-SDN)	1,210	7.879	0.525	0.521	0.652
Seeing at a Distance (SDN)	1,332	7.952	0.513	0.508	0.636

Note. READFKGL = Flesch-Kincaid grade level (0–12); LSAassa = latent semantic analyses, sentence to sentence, adjacent, mean; LSApssa = latent semantic analyses, sentences, all combinations, mean; LSAppa = latent semantic analyses, paragraph to paragraph, mean; SDN = scientific discovery narrative.

that the attempt to control engagement and interest across the narrative and expository texts was successful.

Measures. The primary dependent variable in this study was the students' understanding and recall of the conceptual information that had been presented in the experimental texts. As there were two topics, there was a need for two measures of understanding: one related to the science content presented in the *Radioactivity* texts and one related to the science content presented in the *Seeing at a Distance* texts. Only content clearly presented in both narrative and expository versions was of interest. For the initial testing, 10 multiple-choice and constructed-response items for each topic were created. A second set of similar items for each topic was created for the second testing session.

Measures of these variables were developed following the framework of the BEAR Assessment System (BAS; Wilson, 2005), which draws upon assessment principles laid out by the National Research Council (NRC; 2001) in the seminal publication *Knowing What Students Know* to guide the development of measures of students' competence in educational domains. The BAS translates the NRC's "assessment triangle" of cognition, observation, and interpretation into a test-construction method that can be summarized in three steps:

1. Develop a model (or *construct map*, in BAS parlance) of student understanding (or knowledge, competence, etc.) in a specified domain area, based on available theory.

2. Develop ways of making and coding observations about the students that provide evidence of their levels of understanding.

3. Link the observations back to the construct map via an appropriate measurement model, such as the Rasch model, and collect and evaluate evidence of the success with which the observations match theory-based expectations.

In the present study, the domains of understanding are restricted to the conceptual information presented in the experimental texts. Kintsch's (1998) construction–integration (C-I) model of reading comprehension, along with experience in science education, was used to develop a construct map of progressively more sophisticated levels of understanding of the conceptual information that students might achieve based on their readings of the texts; this model was then used to help guide the writing of test items intended to be progressively more difficult. This procedure helped generate a test on which different levels of performance could be meaningfully interpreted as reflecting not just higher or lower total levels of knowledge of discrete facts, but also more or less sophisticated levels of comprehension of the relevant material. It should be noted that this adaptation of Kintsch's C-I model is meant neither to provide a generalizable framework for representing scientific conceptual understanding nor to imply that within-student learning progressions should necessarily be viewed in this manner.

The simplest items involved direct recall of factual information that had been explicitly stated in the texts (what Kintsch,1998, refers to as micropropositions; e.g., "chlorine is a type of element"). Somewhat more sophisticated are inferences regarding factual information implied but not explicitly stated in the texts (macro-propositions; e.g., "the surface of the moon is similar to the surface of the earth," which is a reasonable inference given the description of the surface of the moon). Above this in sophistication is understanding of key conceptual "big ideas" of the text (textbase macrostructure; e.g., "learning about the moons of Jupiter helped us understand that the Earth is not the center of the universe"). Finally, there is application of integrated knowledge in novel situations (the situation model; e.g., "how can you show that a glow stick is not radioactive?"). Figure 1 presents the construct map of progressively more sophisticated levels of understanding of the text-based information.

Both multiple-choice and constructed-response items for the two testing sessions were developed to target each of the levels described in the construct map for both topics. Table 2 presents samples for both sessions.

Quality control. The measures went through the same process of editing as the experimental texts. The aforementioned expert panels gave feedback on the content, format and wording of each item. The 15 students involved in the pilot study also completed the measures in "think-aloud" fashion; their verbal and written responses were analyzed to ensure that it was variation in understanding of the targeted domain, and not variation in construct-irrelevant factors such as the particular wording of the items, that was giving rise to variation in item responses, and to guide further revisions when this was not the case.

After edits and the removal of some test questions based on initial pilot testing, a total of 16 questions were used for the *Radioactivity* measure, and a total of 17 questions were used for the *Seeing at a Distance* measure. These questions were divided into two forms each, with approximately half of the items on each form.

Scoring guides for open-ended questions were developed, and the authors of this study and a research assistant served as scorers. Approximately half of the responses were blindly scored by at least two raters. After initial training sessions, interrater reliability (as estimated by Fleiss's kappa, with a simple dichotomous scoring of 1 if raters agreed and 0 if they did not) of assigned scores exceeded .90. Remaining differences in scoring were resolved through discussion and consensus.

Procedure

Data collection occurred during the participants' science class period, in the presence of the teacher. Students were randomly assigned to one of four conditions[1]; in each condition, students received the SDN version of one topic (*Radioactivity* or *Seeing at a Distance*) and the expository version of the other. The order in which texts were given was counterbalanced in a modified Latin-square design (Topic $\times$ Genre $\times$ Order). The booklets themselves were subtly color-coded to allow the researchers to keep track of the condition. Following a silent reading session, each text was collected by the teachers, and the students were asked to complete the appropriate outcome measure (i.e., the items related to the topic

[1] To ensure the success of the randomization, the number of students of each gender and ethnicity and mean levels of English language-arts proficiency were compared across all four groups. No significant differences in any of these variables were observed.

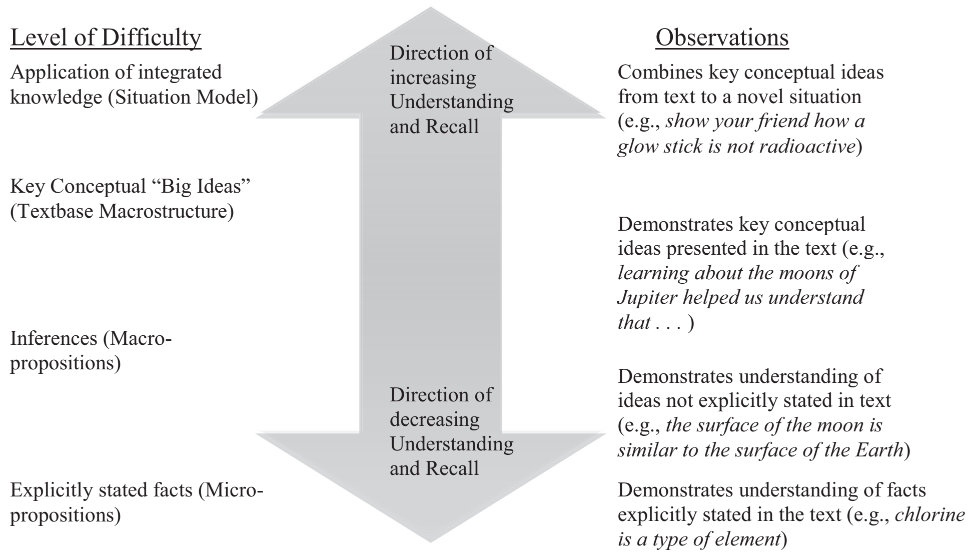

Figure 1. Construct map of understanding and recall of experimental texts.

of the text). Students were also asked which of the texts they preferred.

Prior to the first session, we visited each participating classroom to introduce the project. To mitigate the potential influence of stereotype threat (Steele & Aronson, 1995), we attempted to introduce the study in a way that encouraged students not to perceive it as an evaluation. Instead, students were told their help was needed to improve the quality of the texts and they were invited to share any feedback they had. The outcome measures were framed as tasks that helped us judge the quality of the texts. Unnecessary technical and academic jargon was avoided.

Teachers were asked to give a ranking of their students by their demonstrated knowledge (classroom performance). The four teachers each provided a ranked list of their students from each of their classes. Teachers confirmed that they mainly used student grades for this determination and then made further judgments about the relative performances of students with the same grades.

One week after the first session, teachers administered the second form of the outcome measures in class. This second phase of items was administered as a minimal check of retention, similar to the procedures of Jetton (1994). The length of time between testing sessions (1 week) was chosen simply to ensure that any differences in demonstrated understanding of text information overcame the initial 24-hr forgetting curve (Baddeley, 1999; Ebbinghaus, 1885/1913). Students did not revisit the texts during the second session.

Results

The primary purpose of the study was to assess whether presenting information encased in a scientific discovery narrative rather than in traditional expositional format led to greater levels of understanding. Investigation of this issue in the context of this study requires that the outcome measures be considered valid measures of understanding. Thus, in this section we briefly present results from analysis of the measures, followed by a comparison of results from the narrative and expository conditions.

Evaluation of Measures

The final step in the test-construction strategy advocated by the BEAR Assessment System is the evaluation of the quality of measures through psychometric models. In the present case, the scales were designed with the intention of measuring unidimensional, continuous latent person variables (understanding of the two specific domains covered by the texts). Thus, the Rasch model (Rasch, 1960/1980) was used for the purposes of both evaluation of scale characteristics and estimation of latent person proficiencies. ConQuest (Version 2.0; Wu, Adams, Wilson, & Haldane, 2007) was used for all analyses.

Items on both the *Radioactivity* and *Seeing at a Distance* measures were analyzed both as a whole and separately for the two forms. All items on both measures save one displayed adequate fit to the Rasch model, with both infit (weighted) and outfit (unweighted) mean squared residual fit statistics falling between 0.8 and 1.2.[2] The sole misfitting item was discarded and is not reflected in the tables and analyses presented here.

Estimates of person-separation reliability (analogous to Cronbach's alpha in true score theory models), along with the number of items on each measure, are given in Table 3. Estimates are given for each measure (*Radioactivity* and *Seeing at a Distance*) when all items are calibrated together and when calibrated separately. The lower apparent reliability of the separate calibrations can be explained by the lower number of items associated with these individual administrations.

After correcting for attenuation due to measurement error, scores on the California state English–Language Arts test (CST-ELA; estimated reliability = .95; Educational Testing Service, 2008) were estimated to correlate with scores on the *Radioactivity* measure at r = .74 and with scores on the *Seeing at a Distance* measure at r = .62.

[2] There are many rules of thumb for interpreting fit statistics; the range given here (suggested by Bond & Fox, 2007, for high-stakes tests) is the most conservative rule of which we are aware.

Table 2
Sample Items for Both Sessions

Radioactivity	*Seeing at a Distance*
Session 1	
3. Not all of the items below are elements. Circle only the elements. [eight substances and elements are listed]	3. Draw and label the parts of the Galilean telescope. [open-ended]
4. Your friend shows you a glow stick and thinks it is radioactive because it glows. Explain at least two ways to show your friend that the glow stick is not radioactive. [open-ended]	4. Use two of the following words to fill in the blanks in the following sentence: "In order to clearly see details on distant objects, the _____ [concave] lens must have a greater curve than the _____ [convex] lens in a telescope."
Session 2	
5. Radioactive elements are different from nonradioactive elements. Write down all the possible ways that make radioactive elements special. [open-ended]	5. Match the following two lenses with the correct label: [drawing of two lenses] (a) concave (b) convex
6. Pitchblende is: (a) an ordinary rock (b) a radioactive element (c) a radioactive rock (d) a uranium element	6. Before Galileo, telescopes could only help us see ships in the ocean 60 miles away. Write down all the possible ways that make the Galilean telescope much more powerful. [open-ended]

Estimated correlations with teacher-ranked performance in science were $r = .37$ for *Radioactivity* and $r = .54$ for *Seeing at a Distance*. (The correlation between the CST-ELA and teacher-ranked science performance was estimated at $r = .50$.) These estimated correlations are consistent with the theory that the scales developed for this study measure domains of academic performance but are not synonymous with overall science performance or knowledge.

Comparison of Conditions

Given the design of the study, and given the assumption that the outcome measures are valid measures of understanding, an effect of condition (narrative vs. nonnarrative) would be apparent in group-level mean differences on the measures.[3] In Table 4, the average raw score[4] for students on each of the four measures is shown for students from each of the two main conditions. As can be seen, there was a statistically significant difference between the SDN and nonnarrative conditions for *Seeing at a Distance* text at the first testing and for both topics at the second testing. The standardized mean differences between the two conditions were noticeably larger at the second testing (as large as one and a half standard deviation units).[5]

Interactions

We also investigated the interactions between the effects of treatment and school, and treatment and grade level, to explore the possibility that exposure to the SDN could exert a stronger influence on performance for certain groups of participants. In this sample, school is confounded with both ethnicity and socioeconomic status, as the vast majority of non-Caucasian (specifically, Latino and African American) students were enrolled at Oaks Middle, which was the more economically disadvantaged of the two schools.

Multiple regression models were fit, with raw scores on the outcome measures regressed on a dummy variable for condition (1 if narrative, 0 if nonnarrative), a dummy variable for school (1 if Ocean Middle, 0 if Oaks Middle) or grade (1 if eighth grade, 0 if seventh), and an interaction term formed by the product of the first two variables.

Table 3
Number of Items and Estimates of Person-Separation Reliability

Measure	# Items	Estimated reliability
Radioactivity, total	16	.82
Radioactivity, Time 1	8	.72
Radioactivity, Time 2	8	.68
Seeing at a Distance, total	17	.79
Seeing at a Distance, Time 1	8	.62
Seeing at a Distance, Time 2	9	.69

Note. Reliability was estimated as Rasch person-separation reliability (directly analogous to Cronbach's alpha in true score theory).

[3] The clustered design of the study (209 students sampled from nine classrooms) indicates that the independence-of-observations assumption of parametric statistical models is probably violated to some extent and that standard errors could negatively biased as a result. To investigate possible problems with such dependencies, we employed one-way random effects analyses of variance to calculate the intraclass correlation coefficients (ICCs) of the Rasch-scaled test scores. It was found that neither of the ICCs (for either the *Seeing at a Distance* or the *Radioactivity* measures) were statistically significantly greater than zero. Thus, within-class dependency is not likely to be a significant source of bias in the results presented in this article.

[4] Raw score was calculated by taking the sum of the score of each item. Dichotomous items were scored 0/1; items on which partial credit was available were scored using fractional values (e.g., 0/.5/1).

[5] The results described here were replicated using latent regression models (Adams, Wilson, & Wu, 1997) via ConQuest. These models can be seen as latent variable extensions of simple t tests, with the difference that latent regression models directly estimate the impact of predictors on a latent variable, thus controlling for measurement error, and avoiding problems with attenuation associated with the two-step procedure of first estimating student scores and then treating such scores as observed variables in a standard statistical model. In the present case, results from these analyses (available from the authors) closely mirrored the presented results from raw-score differences; the same pattern of statistical significance was observed, and effect sizes were similar to a margin of error of approximately 10%.

Table 4
Average Raw Score of Students in Each Condition

Measure	Narrative version		Difference	Standardized difference
	Seeing at a Distance *M* (*SD*)	*Radioactivity* *M* (*SD*)	*M* (*SE*)	(effect size)
Radioactivity, Time 1	5.85 (2.39)	6.21 (1.93)	.37 (.30)	.17
Radioactivity, Time 2	4.32 (2.12)	5.15 (1.75)	.84** (.14)	.42
Seeing at a Distance, Time 1	6.54 (2.01)	4.52 (2.19)	2.02** (.29)	.96
Seeing at a Distance, Time 2	7.00 (1.98)	4.17 (1.82)	2.83** (.20)	1.49

Note. The sample size was $N = 209$ at Time 1, and $N = 192$ at Time 2.
** $p < .01$.

As can be seen in Tables 5 and 6 findings revealed significant interactions between both condition and school and condition and grade. The models shown in Table 5, which investigate the interaction between treatment and school, show that the effect of treatment was significantly higher for the students at Oaks Middle School (the higher poverty, more predominantly Latino and African American school) for both of the *Radioactivity* and the first of the *Seeing at a Distance* outcomes.

As is shown in Table 6, there was a negative Grade × Treatment interaction for the second testing of *Radioactivity*. This effect (–.66) is of approximately the same magnitude as the main effect of treatment (.64); this can be interpreted as indicating that there were no effects of treatment for either *Radioactivity* outcome for the eighth grade participants. The seventh grade participants, however, did benefit from the SDN version for this topic.

Discussion

Results from this study provide evidence in support of the hypothesis that middle school students benefit from reading texts that highlight the process of discovery. This appeared to be especially true for the seventh grade students and the students at the more economically disadvantaged, predominantly non-Caucasian school. The addition of narrative qualities to otherwise matched texts allowed for a direct test of the hypothesis that such qualities would serve as a distraction rather than a supportive element to the reader. Thus, this study sought to investigate the importance of highlighting the emotional values and intentions of scientists in scientific texts.

The fact that the narrative versions were longer than their expository counterparts could be considered a limitation to the study, as it could be argued that students in the narrative conditions simply spent more time on task than their nonnarrative counterparts. Here we would note that the purpose of the study was precisely to determine whether the inclusion of narrative elements does constitute time on task, rather than serving as an irrelevant or negative distraction.

Results suggest that the SDN supports retention of science conceptual information beyond the initial 24-hr forgetting curve. Further, the interactions between treatment and school suggest that students from non-Caucasian or less socioeconomically advantaged backgrounds benefit more from reading science texts that highlight discovery. However, it is important to acknowledge two significant limitations in this analysis. First, the design of this study and the resulting number of non-Caucasian participants was not great enough to allow for anything more than the crudest of comparisons of cultural affiliation. Further, it was not possible to separate ethnic minority status and socioeconomic status for this study. Future research could address this confound.

The interaction between grade and treatment for the *Radioactivity* texts indicates that seventh grade participants benefited significantly more than eighth grade participants from receiving the narrative text version. One possible explanation for this may be found in the fact that, as discussed previously, many of the eighth grade students had recently finished a school unit related to chemical reactions and could be presumed to possess higher levels of

Table 5
Standardized Regression Coefficients of Outcome Measures on Treatment, School, and Their Interaction

Outcome	Treatment (SDN/non-SDN)	School (Ocean = 0, Oaks = 1)	School × Treatment
Recall I *Radioactivity*	0.02 (.06)	−1.18 (.19)**	.55 (25)*
Recall II *Radioactivity*	0.08 (.13)	−1.18 (.17)**	.91 (.25)**
Recall I *Seeing at a Distance*	0.67 (.14)**	−0.42 (.18)**	.52 (.26)*
Recall II *Seeing at a Distance*	1.20 (.13)**	−0.36 (.17)**	.36 (.25)

Note. Standard errors are in parentheses. The sample size was $N = 209$ at Time 1, and $N = 192$ at Time 2. SDN = scientific discovery narrative.
* $p < .05$. ** $p < .01$.

Table 6
Standardized Regression Coefficients of Outcome Measures on Treatment, Grade, and Their Interaction

Outcome	Treatment (SDN/non-SDN)	Grade Level (7th grade = 0, 8th grade = 1)	Grade × Treatment
Radioactivity, Time 1	0.21 (.18)	0.71 (.20)**	0.04 (.28)
Radioactivity, Time 2	0.64 (.15)**	0.93 (.16)**	−0.66 (.24)**
Seeing at a Distance, Time 1	0.79 (.16)**	0.46 (.17)**	−0.04 (.25)
Seeing at a Distance, Time 1	1.35 (.15)**	0.34 (.16)**	0.13 (.23)

Note. Standard errors are in parentheses. The sample size was $N = 209$ at Time 1, and $N = 192$ at Time 2. SDN = scientific discovery narrative.
** $p < .01$.

relevant knowledge prior to the experiment; such students would presumably do well regardless of genre of text. Differences in prior knowledge could also at least partially account for the Treatment × School interaction, as students in the more economically advantaged school could also be expected to be more likely to do well regardless of treatment. However, as individual levels of prior knowledge were not assessed in this study, these explanations remain speculative.

Overall, students indicated a preference for the *Radioactivity* text over the *Seeing at a Distance* text, suggesting that the former text was more inherently interesting, regardless of genre. The greater differences between the outcomes for the narrative and nonnarrative versions of the *Seeing at a Distance* text may then be due in part to the fact that the topic of radioactivity was more engaging to begin with, and thus the narrative encasement did not matter as much. In particular, even the nonnarrative version of the *Radioactivity* text included a discussion of mortality, a very fundamentally human concern, though it did so in general terms rather than through the personal examples of Marie Curie and her husband.

The precise cognitive and affective mechanisms through which presentation of information in the form of a discovery narrative benefits learning are still not fully clear. However, empirical findings on the personalization principle and visible author converge on the idea that personal reader-to-text connections result in greater attentiveness to conceptual content. We suggest that the SDN exposes the readers even more to the humanness of science, which encourages greater invested attention on the part of students, in that the readers have the opportunity to vicariously experience the scientific journey of discovery. This increased interest and attention can facilitate deeper understanding and recall of the pertinent information.

SDNs highlight the process of discovery, and as such may be even more useful when the explicit goal of instruction is to teach about the process of science in addition to or instead of its content. We note that this point is not addressed directly in this study; the outcome measures were constructed carefully to assess only knowledge that could be gained from either version of each text, and therefore do not directly assess the student's understanding of the scientific process. Nonetheless, SDNs may be valuable in their own right in that they bring more focus to the process of discovery of scientific information, especially in light of the empirical findings that suggest a lack of understanding about the process of acquiring and constructing scientific knowledge (i.e., nature of science as a way of knowing) among elementary students and teachers alike (Akerson & Abd-El-Khalick, 2005; Gallagher, 1991; Griffiths & Barman, 1995; Lederman, 1992; Moss, 2001). If teaching students about the process of science (in addition to its content) is a goal, the SDN will surely be of even more value than these results illustrate.

We noted in the introduction to this article that students and teachers alike often have little knowledge about science as a way of learning something about our world. We also noted different strands of scholarship both in favor of and against the inclusion of narratives in education. In this study, we hope to have contributed to this discussion by showing how a particular form of narrative—one that presents science as a human process of discovery—can lead to greater levels of conceptual understanding and knowledge retention. We hope future investigations involving issues of genre consider this broader perspective on narrative.

References

Adams, R. J., Wilson, M. R., & Wu, M. L. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22*(1), 47–76.

Akerson, V. L., & Abd-El-Khalick, F. S. (2005). "How should I know what scientists do? I'm just a kid": Fourth-grade students' conceptions of nature of science. *Journal of Elementary Science Education, 17*(1), 1–11. doi:10.1007/BF03174669

Allchin, D. (2003). Scientific myth-conceptions. *Science Education, 87,* 329–351. doi:10.1002/sce.10055

Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education, 31,* 1683–1707. doi:10.1080/09500690802380695

Baddeley, A. D. (1999). *Essentials of human memory.* Hove, England: Psychology Press.

Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). London, England: Psychology Press.

Bowker, G. C. (2005). *Memory practices in the sciences.* Cambridge, MA: MIT Press.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

California State Board of Education. (1998). *Science content standards for California public schools.* Retrieved from http://www.cde.ca.gov/be/st/ss/documents/sciencestnd.pdf

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Conant, J. B. (1957). *Harvard case histories in experimental science.* Cambridge, MA: Harvard University Press.

Curie, M. (1904). *Radioactive substances.* London, England: Chemical News Office.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology.* New York, NY: Teacher's College Press.

Educational Testing Service. (2008). *California standards tests: Technical report: Spring 2007 administration*. Retrieved from California Department of Education website, http://www.cde.ca.gov/ta/tg/sr/documents/csttechrpt08.pdf

Gallagher, J. J. (1991). Prospective and practicing secondary school science teachers' knowledge and beliefs about the philosophy of science. *Science Education, 75,* 121–133. doi:10.1002/sce.3730750111

Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). "Seductive details" and learning from text. In A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 239–254). Hillsdale, NJ: Erlbaum.

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6,* 41–57.

Glynn, S. M., & Muth, K. D. (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of Research in Science Teaching, 31,* 1057–1073. doi:10.1002/tea.3660310915

Greenfield, P. M. (1997). Culture as process: Empirical methods for cultural psychology. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology: Vol. 1. Theory and method* (2nd ed., pp. 301–346). Needham Heights, MA: Allyn & Bacon.

Griffiths, A. K., & Barman, C. R. (1995). High school students' views about the nature of science: Results from three countries. *School Science and Mathematics, 95,* 248–255. doi:10.1111/j.1949-8594.1995.tb15775.x

Hidi, S., Baird, W., & Hildyard, A. (1982). That's important but is it interesting? Two factors in text processing. In A. Flammer & W. Kintsch (Eds.), *Advances in psychology: Discourse processing* (pp. 63–75). New York, NY: North-Holland. doi:10.1016/S0166-4115(08)62681-3

Jetton, T. L. (1994). Information-driven versus story-driven: What children remember when they are reading informational stories. *Reading Psychology, 15(2),* 109–130. doi:10.1080/0270271940150203

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology, 97,* 235–245.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 23,* 1567–1577.

Kintsch, W. (1998). *Comprehension.* Cambridge, England: Cambridge University Press.

Klopfer, L. E., & Cooley, W. W. (1963). The history of science cases for high schools in the development of student understanding of science and scientists. *Journal of Research in Science Teaching, 1,* 33–47. doi:10.1002/tea.3660010112

Koch, A., & Eckstein, S. G. (1995). Skills needed for reading comprehension of physics texts and their relation to problem-solving ability. *Journal of Research in Science Teaching, 32,* 613–628. doi:10.1002/tea.3660320607

Lassen, S. R. (2007). Impact of school-wide PBS on indicators of social development and academic performance in an inner-city middle school. *Dissertation Abstracts International,* 67(11). (UMI No. 3243448)

Latour, B., & Strum, S. (1986). Human social origins: Oh please, tell us another story. *Journal of Social and Biological Structures, 9,* 169–187. doi:10.1016/0140-1750(86)90027-8

Lederman, N. G. (1992). Students' and teachers' conceptions about the nature of science: A review of the research. *Journal of Research in Science Teaching, 29,* 331–359. doi:10.1002/tea.3660290404

Leite, L. (2002). History of science in science education: Development and validation of a checklist for analyzing the historical content of science textbooks. *Science & Education, 11,* 333–359. doi:10.1023/A:1016063432662

Lemke, J. (1990). *Talking science: Language, learning and values.* Westport, CT: Ablex.

Lipka, J. (1991). Toward a culturally based pedagogy: A case of one Yup'ik Eskimo teacher. *Anthropology and Education Quarterly, 22,* 203–223. doi:10.1525/aeq.1991.22.3.05x1050j

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education: Theory and Practice, 12,* 73–84. doi:10.1080/13603120701576241

Madrazo, G. M. (1997). Using trade books to teach and learn science. *Science and Children, 34,* 20–21.

Mayer, D. A. (1995). How can we best use children's literature in teaching science concepts? *Science and Children, 32,* 16–19, 43.

Mayer, D. A. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

McLaren, B. M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006, June). *Studying the effects of personalized language and worked examples on the context of a web-based intelligent tutor.* Paper presented at the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005). Coh-Metrix (Version 1.4) [Computer software]. Retrieved January 20, 2010 from http://cohmetrix.memphis.edu

Metz, D., Klassen, S., McMillan, B., Clough, M., & Olson, J. (2007). Building a foundation for the use of historical narratives. *Science & Education, 16,* 313–334. doi:10.1007/s11191-006-9024-z

Mishler, E. G. (1995). Models of narrative analysis: A typology. *Journal of Narrative & Life History, 5*(2), 87–123.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358–368. doi:10.1037/0022-0663.91.2.358

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality environments: Role of method and media. *Journal of Educational Psychology, 94,* 598–610. doi:10.1037/0022-0663.94.3.598

Moss, D. M. (2001). Examining student conceptions of the nature of science. *International Journal of Science Education, 23,* 771–790. doi:10.1080/09500690010016030

Murata, K. (2007). Unanswered questions: Cultural assumptions in text interpretation. *International Journal of Applied Linguistics, 17*(1), 38–59. doi:10.1111/j.1473-4192.2007.00132.x

National Committee on Science Education Standards and Assessment & National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press. Retrieved from http://www.nap.edu/openbook.php?isbn=0309053269&page=160

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

Niaz, M., & Rodriguez, M. A. (2000). Teaching chemistry as rhetoric of conclusions or heuristic principles: A history and philosophy of science perspective. *Chemistry Education: Research and Practice in Europe, 1,* 315–322.

Nolen, S. B. (1995). Effects of a visible author in statistical texts. *Journal of Educational Psychology, 87,* 47–65.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87,* 224–240. doi:10.1002/sce.10066

Paxton, R. J. (1997). "Someone with a life wrote it": The effects of visible author on high school history students. *Journal of Educational Psychology, 89,* 235–250. doi:10.1037/0022-0663.89.2.235

Paxton, R. J. (2002). The influence of author visibility on high school students solving a historical problem. *Cognition and Instruction, 20,* 197–248. doi:10.1207/S1532690XCI2002_3

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: The University of Chicago Press. (Original work published 1960)

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology, 80,* 16–20.

Reeves, B., & Nass, C. (1996). *The media equation.* Cambridge, England: Cambridge University Press.

Rice, D. C. (2002). Using trade books in teaching elementary science: Facts and fallacies. *The Reading Teacher, 55,* 552–565.

Ricoeur, P. (1980). Narrative time. *Critical Inquiry, 7*(1), 169–190. doi:10.1086/448093

Simmons, R., & Blyth, D. (1987). *Moving into adolescence: The impact of pubertal change and the school context.* New York, NY: Aldine de Gruyeter.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328,* 450–452. doi:10.1126/science.1182597

Solomon, J., Duveen, J., Scot, L., & McCarthy, S. (1992). Teaching about the nature of science through history: Action research in the classroom. *Journal of Research in Science Teaching, 29,* 409–421. doi:10.1002/tea.3660290408

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811. doi:10.1037/0022-3514.69.5.797

Stewart, E. B. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school- and individual-level factors on academic achievement. *Education and Urban Society, 40,* 179–204. doi:10.1177/0013124507304167

van Dijk, T. A. (2001). Critical discourse analysis. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 352–371). Oxford, England: Blackwell.

Wade, S. E., & Adams, R. B. (1990). Effects of importance and interest on recall of biographical text. *Journal of Reading Behavior, 22,* 331–353.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1997). Learning influences. In H. J. Walberg & G. D. Haertel (Eds.), *Psychology and educational practice* (pp. 199–211). Berkley, CA: McCuthan.
Welsh, W. N. (2000). The effects of school climate on school disorder. *Annals of the American Academy of Political and Social Science, 567*(1), 88–107. doi:10.1177/0002716200567001007
Westerlund, J., & Fairbanks, D. (2004). Gregor Mendel and "myth-conceptions." *Science Education, 88,* 754–758. doi:10.1002/sce.20007
White, H. (1980). The value of narrativity in the representation of reality. *Narrative Inquiry, 7*(1), 5–27.
Widdowson, H. G. (2004). *Text, context, pretext: Critical issues in discourse analysis.* Oxford, England: Blackwell. doi:10.1002/9780470758427
Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Erlbaum.
Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. J. (2007). ACER ConQuest (Version 2.0) [Computer software]. Hawthorn, Australia: ACER.

Received October 4, 2010
Revision received February 27, 2012
Accepted March 13, 2012 ■