Education Tech Research Dev (2015) 63:517–537
DOI 10.1007/s11423-015-9390-9

RESEARCH ARTICLE

Contrasting case instruction can improve self-assessment of writing

Xiaodong Lin-Siegler[1] · David Shaenfield[2] · Anastasia D. Elder[3]

Published online: 20 June 2015

Abstract Self-assessment is a process during which students evaluate the quality of their work in a given domain based on explicitly stated criteria. Accurate self-assessments improve students' academic achievement. Yet, students often have difficulties assessing their own work. It is possible that appropriate instructional supports will help students overcome these difficulties. To test this premise, we compared the effects of presenting and discussing examples of well and poorly written stories (contrasting cases) with the effects of only presenting and discussing examples of well written stories (good cases only) on students' writing. Fifty-three 6th-grade students in two history classrooms were randomly assigned to either the contrasting cases or good-cases-only instructional conditions. Results showed that students in the contrasting cases instructional condition created stories of better quality, developed a deeper understanding of the assessment criteria, and became better able to identify areas in need of improvement. This study is one of few efforts applying perceptual learning theories to improve academic skills in everyday classroom settings. The use of contrasting cases provides a promising yet a simple instructional approach that both teachers and students can use to improve writing and self-assessment.

Keywords Self-assessment · Contrasting cases · History story writing

Introduction

Self-assessment is a learning process during which students evaluate the quality of their work by comparing it to explicitly stated criteria and making needed revisions (Andrade 2010; Andrade and Warner 2012). It is an important component of students' self-regulated

✉ Xiaodong Lin-Siegler
xlin@tc.columbia.edu

[1] Teachers College, Columbia University, New York, NY 10027, USA

[2] Sacred Heart University, Fairfield, CT 06825, USA

[3] Mississippi State University, Starkville, USA

learning processes and academic success (Dunlosky and Rawson 2012; Kostons et al. 2011; Pintrich 2004; Zimmerman 2006, 2008). Self-regulation, especially self-assessment is often difficult for students, especially young and poor performing students (Azevedo et al. 2008; Elder 2010; Kostons et al. 2011). One possible explanation is that this may be due to a lack of appropriate instructional support. To investigate this premise, we conducted a classroom-based experiment to test whether contrasting-case based instruction that engages students in analyzing and using contrasting examples (good vs. poor) while studying assessment criteria as opposed to the use of good examples only would improve the quality of the work students produce and the self-assessment of that work.

Two lines of existing theories and research provide support to our study. The first line involves theories and research suggesting that self-assessment is a critical component of self-regulated learning processes, and it facilitates self-regulation of writing. The second line of research concerns the potential benefits of instructional support on the development of students' self-assessment abilities. In particular, why contrasting case based instruction is believed to help students develop these abilities.

Self-assessment is a critical component of self-regulated learning processes

Self-assessment is a critical component of self-regulated learning (Azevedo et al. 2008; Falchikov and Boud 1989; Greene and Azevedo 2007; Kostons et al. 2011; Zimmerman and Schunk 2011). Self-regulation process consists of three phases: planning, execution, and revision (Greene and Azevedo 2007; Kitzantas and Zimmerman 2006; Winne 2005; Zimmerman 2000). These phases interact in a cyclical process in which the students plan their learning tasks, perform the tasks, monitor during performance, evaluate the results, and make revisions (Labuhn et al. 2010; Zimmerman 2000; Zimmerman and Kitsantas 2002). Self-assessment takes place throughout the regulation cycle. As such, self-assessment, regulation and learning together enable students to achieve a desired learning outcome (De Bruin and Van Gog 2012; Dunlosky and Hertzog 1998; Thiede and Dunlosky 1999). For example, when students used judgments of how well (self-assessment) they understood certain concepts to study for their upcoming exam, they typically spent more time studying the concepts that they judged that they did not know well (self-regulation of learning) (Higham 2013; Metcalfe 2009; Van Loon et al. 2013).

Studies of self-regulated learning typically asked students to select which part of the learning materials they wish to re-study or reanalyze, before or after taking tests (De Bruin and Van Gog 2012). The correlation between students' assessment of what they should re-study and what they actually study served as an indication of the quality of their self-regulation. Accurate assessments helped students make decisions over what part of their work needed improvement and where they should invest more effort (Koriat 2012; Koriat et al. 2006; Lin 2001; Nelson 1984). Students who were able to assess their own learning usually comprehended instruction better and solved problems more effectively (Brown 1978, 1987; Flavell and Wellman 1977; Pintrich et al. 2003; Schraw 2009; Veenman 2011; Winne 2005; Zimmerman 2006, 2008). The focus of the present study is to help students develop self-assessment skills in order to improve their self-regulation and academic performance.

Self-assessment facilitates self-regulation of writing

Writing is a pervasive activity that is crucial for students' success in schools (Graham and Hebert 2011; Graham et al. 2012; Scardamalia and Bereiter 1985). Flower and Hayes (1980) argued that writing is a self-regulated learning process. It involves planning what to write, gathering information relevant to the main writing topic, organizing thoughts, generating sentences, self-assessing if writing goals are met and evaluating the quality of the completed products. "This process is not a matter of simply having students determining their own grades or a rating…" (Andrade et al. 2010, p.199). Students need to reflect upon and articulate the strengths and weaknesses in their writing and identify specific areas that need improvements in order to produce high quality work (Andrade 2001, 2010; Andrade et al. 2008; Falchikov and Boud 1989; Eva and Regehr 2008a, b; Sargeant et al. 2008).

Thus, self-assessment of one's own writing involves integrating one's own writing with information from external sources, such as criteria and examples for what makes good or poor quality writing (Andrade 2010; Elder 2010; Sargeant 2008). Research by Hacker et al. (2009) found that self-assessment activities improved essay writing in elementary and middle school classrooms. Their findings suggest that training students to use the information obtained from their self-assessments improved not only their subsequent revision of the writing, but other cognitive, affective and social processes involved in writing. Other research studies also found that effective self-assessments led to strategic adjustments in writing behavior (Harris and Graham 1992; Pressley and Harris 2006; Zimmerman and Risemberg 1997).

A considerable body of research evidence suggest that self-assessment of one's own writing is an intrinsically difficult task, especially for struggling young writers (see Andrade and Boulay 2003; Glaser and Brunstein 2007; Graham 2006; Graham and Harris 2003; Harris et al. 2010). This is because they often (1) lack the knowledge to differentiate the characteristics of well-constructed composition from poor ones (Graham and Harris 2000; Harris et al. 2010); (2) show almost no evidence of planning and self-assessment of their writing unless they are explicitly instructed to do so (Graham 2006; Graham and Harris 2000); and (3) have low expectations for their academic work and hold on to inflated judgments of their own writing (Harris et al. 2009; Lin et al. 2010). Yet, teachers often ask their students to engage in self-assessment, when they may not have the skill to do so. Hence, instructional supports should be provided to scaffold students to effectively engage in activities of self-assessment (Pintrich 2000; Zimmerman and Schunk 2011).

Instruction that supports the development of self-assessment

The second line of research relevant to the present study concerns the effects of instructional support for students' self-assessment. One approach to support self-assessment involves teaching students various types of self-regulation and self-assessment strategies during the writing process. For instance, the well-known self-regulated strategy development (SRSD) program (Graham and Harris 2000; Harris et al. 2010) provides multifaceted instructional interventions: teaching both genre-specific strategies for composition (e.g., strategies for setting writing goals and revising stories) in conjunction with multiple self-assessment and self-regulation strategies (e.g., self-instruction, goal setting, assessing work using a set of criteria and examples, and self-revision). In general, the SRSD instructional

model improved writing and self-assessment with above average students more than with struggling students (Graham and Harris 2003). This might be because teaching effective self-assessment strategies did not explicitly address the difficulty associated with a lack of ability to differentiate good quality writing from poor quality writing. Even if a student learned effective self-assessment strategies, the flaws in writing products were often not visible to struggling writers (Glaser and Brunstein 2007). As such, they simply do not know what specific parts of their composition need revision.

Rubrics-based instruction is a tool that teachers often use to scaffold students' self-assessment in writing (Andrade 2001, 2010; Moskal 2003). A rubric usually has two characteristics: a list of criteria for what counts as excellent and poor quality of work; and examples illustrating how the work should look like based on the criteria, so that students can compare their work with the desired examples (Andrade 2001, 2010). Teachers use rubrics as a means of communicating expectations for an assignment, providing focused feedback on work in progress, and grading final products (e.g. Andrade 2001, 2010; Andrade and Du 2005). The process usually begins by teachers and students studying a list of criteria, viewing a good example of a particular assignment, and discussing how various criteria are reflected in the example. Next, students complete the assignment and self-assess their work using the rubric, check list, and example. Finally, students identify areas that they need to improve in their work (Andrade and Warner 2012).

It is often taken for granted that rubrics are an adequate instructional technique for facilitating students' self-assessment of their own writing. Well-developed rubrics have improved the quality of students' writing and knowledge about what counts as effective writing (Andrade et al. 2009). However, depending on the quality of criteria and the examples used in the rubrics, the effects vary. In some studies, students found that rubrics were often abstract and not straightforward enough for them to use, Thus, researchers find it difficult to draw solid conclusions about students' improvement in writing and self-assessment in relation to the use of rubrics (Andrade et al. 2008, 2009; Andrade and Boulay 2003; Norcini 2003; Winne and Nesbit 2010).

Using examples along with criteria was more helpful than providing criteria or examples alone (Andrade et al. 2008). However, simply handing out and explaining a rubric with good examples only could increase students' knowledge about the criteria, but was not effective when students assessed their own writing (Andrade and Boulay 2003). It did not actively engage them in noticing and analyzing distinctive features that differentiate good and poor writing and applying what they learn about these features in assessing their own work (Andrade et al. 2008; Graham and Perin 2007). These findings suggest that, instead, students were more likely to notice important features that differentiate good from poor writing when using *different* examples to exemplify the assessment criteria (Andrade et al. 2008, 2010; Winne and Nesbit 2010).

Contrasting case based instruction

Contrasting cases are instructional materials designed to help students notice distinctive characteristics that they might otherwise overlook (Schwartz et al. 2011; Schwartz and Martin 2004). Contrasting cases can make new properties and features of a given concept explicit so that even novice learners will not miss them (Schwartz et al. 2011). This approach originated in theories of perceptual learning that emphasized people's ability to differentiate knowledge they acquire (Bransford et al. 1989; Gibson 1969; Gibson and Gibson 1955; Schwartz and Bransford 1998). The overall goal of using contrasting cases is to highlight similarities and differences along a common dimension and help people notice

specific dimensions that make the concepts distinctive. This kind of instructional support should be particularly important to struggling writers since they usually have difficulties identifying limitations of their own writing.

Although studies that empirically tested the effects of contrasting case-based instruction on self-assessment of writing are scarce, a number of studies have documented benefits of having students analyze and discuss contrasting examples when learning new subject matter (Gentner et al. 2011; Wang and Baillargeon 2008). For example, contrasting case-based instruction improved school age children's learning mathematical concepts (Hattikudur and Alibali 2010; Richland and McDonough 2010; Rittle-Johnson and Star 2009); children's acquisition of verbal meaning (Childers 2008; Childers and Paik 2009); physics (Hestenes 1987; VanLehn and Van De Sande 2009); social skills (Gick and Holyoak 1983; Thompson et al. 2000) and college students' business analysis abilities (Gentner et al. 2003).

Many researchers noted the importance of presenting the contrasting examples side by side in order to notice relevant distinctions. Gentner et al. (2003), for instance, advocated that analyzing contrasting cases concurrently, rather than one at a time, was key to producing benefits. This was because when cases were examined one at a time, students tended to focus on surface features, had more difficulties in retrieving what was learned, and were less likely to notice important differences between the cases. For example, college students who compared two business cases by reflecting on their similarities and differences concurrently generated higher quality business solution strategies than those students who read and reflected on the same set of contrasting cases sequentially (Gentner et al. 2003).

These findings have not been empirically tested in classroom settings. The experiences of analyzing contrasting examples should help students notice and detect weaknesses in specific areas of their writing for further improvement. Hence, contrasting case instruction should engage students in analyzing how specific components of the criteria are well or poorly implemented in story writing, which in turn should lead to a deeper understanding of the criteria, improved writing, and subsequent self-assessment. This is especially relevant for poor performing students who have difficulties with writing and self-regulation.

In the present study, we investigated whether analyzing and discussing a well written and poorly written story side-by-side (contrasting cases) produced better quality writing and self-assessment than analyzing and discussing two well-written stories side-by-side. We used a pretest, posttest quasi-experimental design in a sixth-grade classroom that served a number of low academically achieving students. Using models of good story writing represented the standard experience in schools and served as our control condition. We hypothesized that students receiving the contrasting cases-based instruction would create stories of better quality, gain deeper understanding of the criteria for assessing their work, and be better able to identify areas in need of improvement, compared to students in more traditional instruction using only good model cases.

Method

Participants

Fifty-three 6th-grade students ($N = 53$) participated in our study. They were from two classes taught by the same social studies teacher at a diverse, public middle school

(65 % African–American, 15 % Hispanic, 15 % Caucasian, and 5 % Asian/Pacific Islander) located in a southern state of the United States. Ninety-two percent of our participants qualified for free lunch and 52 % were female. About 46 % of the participants in this school pass state standardized tests in mathematics and language arts each year.

Design and procedures

Our study lasted 3 days and it took place during the 3 weeks when students studied a unit on Ancient China. In this unit, all of the students researched two different ancient Chinese dynasties, Qing and Ming Dynasties, and wrote a story about a day in a child's life for each of the two dynasties. A week before the intervention, all of the students (1) researched both dynasties, (2) were given six criteria for what makes a good story to guide their story writing (the explanation for the criteria development will be discussed in a later section of the paper), and (3) wrote a story about Qing dynasty (pre-test story). After the students submitted their stories about Qing dynasty and prior to receiving feedback and a grade from their teacher, the classes were randomly assigned to one of the two intervention conditions: (1) the contrasting cases instructional condition ($N = 27$) which received two exemplar stories that contrasted good and poor features of story writing according to the given rubrics; and (2) the good-cases-only instructional condition ($N = 26$) where only the features of good stories were presented in two story examples.

On the first day of the intervention, for both conditions, the teacher began the class by introducing the goals and plans for the class. She then led the class in using the criteria to analyze and discuss the two exemplar stories (either two contrasting stories or two good stories). The analysis and discussions for both conditions centered on what the students like or dislike about the stories, what was interesting about these stories, and the kinds of things that were good and poor in each story. The teacher also asked students to identify specific dimensions of the stories that illustrated how each of the six rubric criteria was implemented in the stories. For instance, the teacher would say: "Rubric criteria #1 says that the stories should have main thesis. Do you think that the example stories have main thesis? If so, how did the author do that? Which sentence(s) or paragraph described the main thesis in the story?"

For both conditions, the teacher started the second day by asking students to write a story about Ming Dynasty (post-test measure). On the third day of the intervention, students in both conditions conducted an assessment of their own stories and submitted a report on what aspects of their stories were particularly strong or weak on the basis of the rubrics, and how they could be improved (see Table 1 for descriptions of instructional activities for both conditions). The two conditions were identical in all instructional activities and the amount of time spent in analyzing the examples. They only differed in the types of example cases, contrasting cases versus good cases only, which were used during instruction.

Table 1 Descriptions of instructional activities during the intervention

Activities for both contrasting and good case only conditions	Examples
Day one: introduced goals of the interventions analyzed and discussed example stories (whole class) students were asked to created a plan for their story writing	"For the next 3 days, we will use double class periods in the morning analyzing and discussing how each of the 6 criteria was used by other students to develop their stories. Analyzing these examples will help you write your story about Ming dynasty and self-assess the quality of your story". So, today, we will analyze and discuss two stories. Tomorrow, you will write your story about Ming dynasty. The day after tomorrow, you will be asked to self-assess the quality of the story. "Are we clear?" The teacher asked similar questions to students in both conditions, but students gave different responses since the two conditions provided different stories (e.g. contrasting vs. good stories only). The teacher asked the following questions throughout the whole class discussions and students took notes and marked all over the stories: "(a) What specific features you like about story A and B? (b) Which stories do you rather read? And why? (c) What these two stories share in common and how are they different? (d) How does story A or B talk about…? Do you like it? Why and why not? (e) Why do you feel that story B is too long and boring? (f) Are there any good features about A and B you would like to combine when you write your story about the Ming dynasty? Why? (g) Are there any parts of story A or B that you absolutely want to avoid doing for your own story writing? (h) Which features of the story A and B reflected the uses each of the 6 criteria? (i) Which parts of the stories you would like to revise because they did not reflect the 6 criteria for what makes a good story?" "Now, please write a plan for how you plan to use each of the criteria in the story about the Ming dynasty that you will write tomorrow. You should give examples for each criteria and how you would use it in your story writing".
Day two: students wrote the story about ming dynasty	"Today, you will use the criteria, the examples we discussed yesterday in class and the plan you created to write your story about a day of a child life in Ming dynasty. Let me know if you run into any questions".
Day three: students assessed their own story	The teacher gave out our self-assessment questionnaire. Students assessed their own stories and other measures were also given out to the students.

Instructional materials

We developed the following instructional materials to help students' writing and the self-assessment of the stories: (1) six criteria (based on rubrics) to guide the students' story writing and the assessment of the stories; (2) two contrasting stories: one illustrating characteristics of a well-written story and the other exemplifying features of a poorly written story; and (3) two well-written stories: highlighting characteristics of what good stories should be like.

Criteria (rubric) development

The teacher and the researchers jointly developed a rubric using criteria for good and poor historical story writing using the Houghton Mifflin Social Studies Textbook Support series (1999) and evidence from research literature (e.g. McCabe and Peterson 1984; Schneider and Winship 2002). Six criteria were recommended as appropriate and frequently used in evaluating quality of story writing in social studies. According to these criteria, a good story should (1) have a clear main thesis explaining what was most important about the ancient time period that the students researched; (2) have detailed examples to explain how people, particularly children, lived their everyday life; (3) make historical facts come alive by including specific characters and events; (4) present events and characters in a logical and connected manner; (5) teach important lessons; and (6) raise some questions about that period of history for further inquiry. We used these criteria to guide our development of the stories used for instructional interventions of the present study (e.g., the two contrasting stories and the two good stories).

Story development

The good story was written with a very compelling main idea. The story was about a girl's life during the period when India was trying to gain independence from Britain. It presented a story about how this girl and her family lived their everyday life during this time. For instance, the story explained why boys usually went to schools while the girls stayed home during that time. The story also used the conversations among family members to make the historical facts and events alive. By referring to a speech by Mohandas Gandhi, the story taught an important lesson: people could have a revolution without using violence and hurting others. The sequence of events and characters were presented in a logical and connected manner. The story ended by raising specific questions that require further research and study, such as: "Will people believe Gandhi and follow his lead? How can we win our independence from Britain without a violent revolution?"

The poor story had the same content and length as the good story, yet it failed to satisfy many of the major criteria listed above. For example, the main thesis of the story was not very clear. It described the girl's life without any further information about the historical contexts when the story occurred. It presented the historical facts and events with no clear logic and connections among them. It was not clear what important lessons one could draw from the story. Only superficial and general questions were raised at the end, such as "what will happen to my country and my family?" Table 2 presents examples of the stories used as contrasting cases in our study.

For the good cases only group, we used the same good Indian story that was used in the contrasting cases condition. An additional story about a different family living in the same

Table 2 Examples of contrasting stories used in the study

An example of a well-written story	This is a story about peace and freedom in India in 1940s. Hello! My name is Sunia. I am 11 years old. I have three older brothers and a younger sister. I live in Tilonia. My village is on the banks of the Ganges River—a good place for farmers like my family to live because there is water even during the dry season. We are not rich people. My parents, aunts and uncles get up very early every morning to get water from the town well and to milk the cows. Then they get the plow ready to go out to the fields. My brothers work during the day and go to school at night, but I do not go to school at all because I am a girl. A girl's proper place is at home, doing domestic work. This is because many people in my village feel that the benefits of a girl's education will be enjoyed by others, since a daughter, typically, leaves her family after marriage. I listen to the grownups talk about the news my uncles bring from the city. One day they saw a very respected man, Mahatma Ghandi. Ghandi is a strong believer in Hinduism. He was giving a speech about kicking the British out of India. People say we would be better off if we ran our own government. Ghandi said, "We can achieve our independence from the British without a war—without weapons and without hurting anyone. Non-violence is our weapon". "That's impossible", my uncle said. "He must be crazy to think the British will give up India without a fight". I hope Gandhi's right about not having a war—it scares me to think about a war right here, in my own village. Ghandi also thinks that we need to make some big changes in our society after we have our own government. He says we should stop discrimination against the untouchables—the families who for centuries have had the nastiest jobs. He says: "They are the Children of God". But other people think the untouchables are only able to do the dirty jobs that no one else wants. I wonder what will happen in my country: Will people believe Ghandi and follow his lead? Why and why not? How can we win our independence from Britain without a violent revolution? Could the untouchables really be treated the same as the rest of us? It's an exciting time to be in India.
An example of a poorly-written story	Hello! My name is Sunia. I am 11 years old. I have three older brothers and a younger sister. I have lots of uncles and aunts. I live in Tilonia, a small village in India. My village is on the banks of the Ganges River—a good place for farmers like my family to live. I stay home helping my parents with household chores. We grow our own food, taking care of cows, etc. The cows are holy animals that cannot be harmed. My parents are very busy every day. Since I stay home most of the time during the day, I also listen to the grownups talk about the news my uncles bring from the city. One day they saw a very respected man, Mahatma Ghandi. He was born and raised in India and went to college in London. He later became a spiritual and political leader in India. He launched a movement of non-violent resistance to the Great Britain's ruling. Gandhi's political and spiritual hold on India was so great that the British rulers dared not to interfere with him. That day, he was giving a speech to a large crowed about his spiritual and political views: India should run itself and should be independent from the British ruling. However, we should earn the independence peacefully, not violently. People say that we would be better off if we ran our own government. I hope we do not have a war—it scares me to think about a war right here, in my own village. My parents, aunts and uncles get up very early every morning to get water from the town well and to milk the cows. Then they get the plow ready to go out to the fields. My uncles go into the city to sell the milk at the market. We kids take care of the cows, water buffaloes, and goats then go out to the jungle to find food. My brothers go to evening school because they have to work in the daytime. I think there is a lot going on in my country right now, but since we live here in the village, we don't hear much about it. The village just got its first radio a few weeks ago. I wonder what will happen to my country and my family.

period time in India was used as an example of a second good story. The contrasting and the good stories were presented to the students to illustrate how these six criteria were exemplified in the context of telling historical stories of India, a unit they have studied previously.

Measures

The following measures were used to assess students' story writing and self-assessment of the writing in both conditions.

Quality of the stories

The two stories students wrote (pre and post intervention) were evaluated for their quality using the aforementioned rubric (six criteria) of good story writing: a good story should (1) have a clear main thesis; (2) offer detailed examples to explain how people, particularly children, lived their everyday life; (3) make historical facts come alive; (4) present events and characters in a logical and connected manner; (5) teach important lessons; and (6) raise questions for further inquiry. The same set of criteria was also used by students to self-assess the stories they produced before and after the intervention.

To minimize bias in grading, the teacher who implemented the contrasting cases instruction intervention and a researcher who was blind to the intervention independently rated how well the stories met each of the six criteria on a 0–5 scale (0—not meeting the criteria at all, 3—somewhat meeting the criteria and 5—completely meeting the criteria). With six criteria, the maximum possible total score for a story is 30 points and the minimum is 0. Cohen's κ was run to determine the level of agreement between the teacher and the researcher ($\kappa = .778$, $p < .0005$) and the disagreements for the rating of each criteria were resolved through discussions.

Depth of understanding of the criteria

The teacher led whole class discussions about how each of the six criteria was exemplified in stories. For instance, the teacher would say: "A good story should have a clear main thesis. Do you think that Story A has a main thesis? If so, which sentence(s) described the main thesis of the story? Ok. What about Story B then? Does it have main thesis?" After the teacher-led whole class discussions, each student was asked to identify examples for how each of the six criteria was implemented in the contrasting stories or the two good stories. The students then wrote down the examples they found in 3×5 index cards, which were then used to guide the story writing and self-assessment of the story. Two researchers independently coded the examples given by the students to determine the number and the types of these examples. For each criterion, we analyzed if the contrasting case condition had an effect on the number of examples students provided for each of the six criteria. In addition, we also analyzed if students gave contrasting examples (e.g., examples of good and poor implementation of the criteria) or good examples only (e.g., examples of good implementation of the criteria).

Criteria with contrasting examples followed this general form: "A good story should have a main thesis. In story A (good story), the story begins by saying: 'this is a story about…", but story B (poorly written story) begins by describing what has happened without saying anything about what this story is about…". Each student was scored 0–6

such that the student received a score of 0 if they did not generate any examples for any of the criteria. They would receive 1 if they included an example for one criterion and 6 if they gave one example for each of the six criteria. In addition, each student was scored dichotomously such that the student received a score of 1 if they included one contrasting example for any of the criteria and a 0 if they did not generate any examples for any of the criteria. This is important because research demonstrates that students tend to focus on the positive aspects rather than the negative aspects of the work they assess (Dunning et al. 2004). Cohen's κ was run to determine the level of agreement between the two coders ($\kappa = .881$, $p < .0005$) and the disagreements for the rating of each criteria were resolved through discussions.

Quality of self-assessment

Students in both conditions were asked to individually assess their stories using a self-assessment worksheet. The quality of students' self-assessment was measured by two kinds of data: (1) their self-assessment of the strengths and weaknesses of the stories; and (2) the accuracy of such assessment. Each student was asked to assess his or her own stories on a self-assessment worksheet. This worksheet included the following questions: (1) What was good about your story? (strengths); (2) What aspects of the story need improvement? (weaknesses); and (3) How could you improve your story? The students specified the strategies they would use to improve their stories.

The responses to each of the three questions were coded as either a substantive or surface level self-assessment. Substantive self-assessment means the student included information regarding any of the six criteria for what makes a good story. Examples of substantive criteria include: "I have good facts about the historical time"; "The way I started off the story was good because I had a main idea"; "I gave a lot of good details about the Tang dynasty"; "My story did not seem to have a main idea;" and "I should write more about the people". Non-substantive means the student focused their self-assessment on surface features of the story writing, such as the length of the story, the spelling mistakes made, or forgetting to use words or phrases. Examples of non-substantive criteria include "It's long"; or "it is good because there is no mistake in spelling". We assigned a score of 1 to substantive self-assessment and 0 to non-substantive assessment.

For the third question regarding the strategies for revising the work, the students again were coded as to whether they responded with a substantive and specific revision (e.g., I need to explain a particular issue more deeply; I will make the story more interesting and rich, or I will add more information to a specific part of the story) or a mechanical/non-specific revision (e.g., I will check the spelling of the writing; I will re-read the story before handing it in). Students were scored dichotomously for answers to each of the three questions: a student received a score of 1 if the self-assessment for that dimension was substantive and a 0 if not. Cohen's κ was run to determine the level of agreement between the two coders ($\kappa = .861$, $p < .0005$) and the disagreements for the rating of each criteria were resolved through discussions. In addition, we also calculated the mean number of strategies (quantity of the strategies) students in each condition generated for improving their story. This number exceeds 1.

With regard to the accuracy of students' self-assessment, we examined if students' perceptions of the quality of their work matched their actual level of quality as judged by the teachers. Correlational analysis was performed to examine the extent to which the students' self-assigned scores matched the scores assigned by the teacher.

Table 3 Mean ratings of the quality of the story

Dimension	Good cases only				Contrasting cases			
	Pre		Post		Pre		Post	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Main thesis	0.54	0.71	1.12	0.94	0.56	0.70	2.89	0.93*
Details of daily family life	2.08	0.89	0.88	1.14	2.00	0.70	1.48	1.40
Historical facts come alive	0.50	0.71	1.04	0.87	0.37	0.56	2.52	1.55*
Presenting in a logical manner	2.19	0.80	1.62	0.50	2.33	0.62	2.93	1.17*
Teaching important lessons	0.31	0.55	0.65	0.69	0.30	0.47	1.48	1.01*
Questions for further research	0.00	0.00	0.04	0.20	0.00	0.00	1.03	1.19*

All groups equal at pre-test on all dimensions

* Significant improvement for contrasting stories ($p < .05$)

Results

Effects on quality of student story writing

Stories written before and after the intervention were compared to assess improvement in writing quality. The quality of each of the two stories was evaluated according to the six criteria for good story writing presented earlier. Repeated-measures ANOVA comparing the two conditions over time were performed for all scores (see Table 3).

Before the intervention, there was no significant difference in the quality of the stories between the two conditions. After the intervention, the stories produced by the contrasting cases students were of significantly higher quality than the stories produced by the good-cases-only students. The contrasting cases students did significantly better than good-cases-only students in five of the six criteria: (1) utilizing a clear main thesis for their stories [$F(1,51) = 41.39$, $p < 0.001$, $\eta p^2 = 0.450$]; (2) making historical facts come alive by including specific characters and events [$F(1,51) = 41.39$, $p < 0.001$, $\eta p^2 = 0.450$]; (3) presenting events and characters in a logical and connected manner [$F(1,51) = 41.39$, $p < 0.001$, $\eta p^2 = 0.450$]; (4) teaching important lessons [$F(1,51) = 41.39$, $p < 0.001$, $\eta p^2 = 0.450$]; and (5) raising some questions about that period of history for future research [$F(1,51) = 41.39$, $p < 0.001$, $\eta p^2 = 0.450$].

Effects on depth of understanding of the criteria

Depth of understanding of criteria was evaluated based on number of and types of examples students provided. To test if the contrasting cases condition had an effect on the number of examples students provided for each of the six criteria, an ANOVA was performed. Good-cases-only students' mean number of examples was 5.08 ($SD = 2.21$) while the contrasting cases students' mean was 5.96 ($SD = 0.19$). This difference was statistically significant [$F(1,51) = 10.398$, $p = 0.043$, $\eta p^2 = 0.783$] suggesting that condition did affect the number of examples provided by students. Those from the contrasting cases condition produced more examples than those from the good-cases-only condition.

The data was further analyzed to determine the content of the examples. The next analysis determined if students gave both good and poor examples, or just good examples. Good-cases-only students' mean number of examples of criteria for what makes a good story was 4.27 ($SD = 2.11$) while the contrasting cases students' mean was 3.00 ($SD = 1.30$). This difference was found to be statistically significant [$F(1,51) = 21.337$, $p = 0.011$, $\eta p^2 = 0.124$] suggesting the students in the good-cases-only group focused significantly more on examples of criteria that makes a well-written story relative to the contrasting cases students. The mean number of examples of criteria for what makes a poor story from the students in the good-cases-only group was 0.81 ($SD = 1.06$) compared to the contrasting cases students' mean of 2.96 ($SD = 1.29$). This difference was also statistically significant [$F(1,51) = 61.527$, $p < 0.001$, $\eta p^2 = 0.467$] suggesting the contrasting cases students focused significantly more on poor story examples of criteria than the students in the good-cases-only group. That is, the contrasting cases students paid more attention to examples of the errors people usually make when writing a story than the correct things people usually do while story writing.

Effects on student self-assessment

We next tested if using contrasting cases leads to an increase in the quality and accuracy of students' self-assessment.

Quality of self-assessment

Quality of self-assessment was evaluated in reference to students' self-appraisal of their stories' strengths, weaknesses and strategies for improvement. Each of these were scored as a 1 or 0 based on whether they attended to the substantive aspects or to surface-level aspects in their self-assessment.

Good-cases-only students' mean score for identifying strengths in their story was 0.65 ($SD = 0.49$) while the contrasting cases students mean was 0.81 ($SD = 0.40$). This difference was not statistically significant [$F(1,51) = 0.343$, $p = 0.191$] suggesting that both groups were capable of noticing strengths in their work. Good-cases-only students' mean score for specifying the weaknesses in their story was 0.35 ($SD = 0.49$). The mean of the contrasting case group was 0.74 ($SD = 0.45$). This difference was found to be statistically significant [$F(1, 51) = 2.062$, $p = 0.003$, $\eta p^2 = 0.157$] which suggests the contrasting cases students were more capable of identifying specific and substantive areas of weaknesses in their stories.

The mean number of different types of strategies generated for improving their story was 0.54 ($SD = 0.90$) for the good-cases-only students while the contrasting cases students mean was 1.56 ($SD = 0.85$). This difference was statistically significant [$F(1,51) = 13.702$, $p < 0.001$, $\eta p^2 = 0.259$] suggesting that in general, the contrasting cases students were more capable of coming up with different types of strategies for improving their story compared to the good-cases-only students.

Good-cases-only students' mean score for generating a substantive strategy for improving their story was 0.31 ($SD = 0.47$) while the contrasting cases students' mean score was 0.67 ($SD = 0.48$). This difference was also statistically significant [$F(1,51) = 1.707$, $p = 0.008$, $\eta p^2 = 0.129$] suggesting that the contrasting cases students were better able to recommend substantive strategies for improving their story than the good-cases-only students. Finally, the score for mentioning a surface level strategy for improving their story was analyzed. Good-cases-only students' mean was 0.54

($SD = 0.51$) while the contrasting cases students mean was 0.41 ($SD = 0.50$). This difference was not statistically significant [$F(1,51) = 0.227$, $p = 0.349$] showing that both groups mentioned surface level strategies, such as making the story longer.

Overall, students in both conditions were able to assess the strengths of their stories. However, in comparison to the students in the good-cases-only condition, the students in the contrasting cases condition were more capable of identifying aspects of their stories that needed improvement and offered substantive strategies for how to go about improving their stories.

Accuracy of self-assessment

Accuracy of students' self-assessment was determined by correlating students' estimation of their overall scores on their story and the actual overall scores they received from the teacher. For the good-cases-only group, a negative correlation was found between the students' scores and the teacher's scores ($r = -0.394$, $N = 26$, $p = 0.047$). However, for the contrasting cases group a positive correlation was found between the students' scores and the teacher's scores ($r = 0.407$, $N = 27$, $p = 0.035$). This suggests that the contrasting cases intervention improved students' accuracy of self-assessment.

Discussion

This study is one of few efforts applying perceptual learning theories to improve academic skills in a classroom setting. It specifically investigates effects of instruction using contrasting cases on middle school students' writing and self-assessment. Results indicate that it is possible to improve middle school students' writing and self-assessment of their writing when providing them with effective instructional support. In comparison to the good-cases-only instructional condition, the contrasting cases instructional condition produced greater improvement in (1) the quality of students' writing, (2) the depth of understanding of the criteria students used for subsequent self-assessments, (3) the quality and accuracy of the assessment of their own writing and (4) the number of different strategies proposed to conduct the revisions.

These findings support and extend previous research that provided strong evidence that multiple, differentiated contrasting cases improved students' understanding and performance in mathematics, physics and business classes (e.g., Rittle-Johnson and Star 2007; Hattikudur and Alibali 2010; Van Lehn and Van De Sande 2009). In our study, the contrasting cases instruction in social study classes aided middle school student writing by encouraging students to notice distinctive features that differentiate good writing from poor writing, which they may miss without such contrasts. In addition to existing successful writing interventions, such as strategy instruction, summarization, and peer assistance (Graham and Perin 2007), and the use of criteria-based rubrics (e.g. Andrade et al. 2010), the use of contrasting examples appears to be a practice that is not only beneficial for students' writing, but also for self-assessment of their writing.

One particularly intriguing finding is that while students in both conditions were able to assess what was good about their stories, and what a good story should be like, the students exposed to the contrasting cases were much better at identifying the weak features of their writing that needed further improvement. This finding not only supports previous work indicating a positive bias in self-assessment (Lin and Bransford 2010; Dunning et al. 2004),

but signifies that obtaining an accurate assessment of what's wrong with one's own work is an intrinsically difficult task, one for which people often do not do spontaneously unless explicit instructional scaffolds are offered (Bjork 1999; Dunning et al. 2004; Tsivitanidou et al. 2011). Analyzing contrasting cases promoted transfer of student understanding of writing criteria: they used the criteria not only in critiquing stories but also transferred that understanding to the task of writing their own story.

We speculate that the contrasting cases provide explicit representations of well-written and poorly written stories. Analyzing and discussing contrasting cases elicited active comparisons and processing of the examples (e.g., Schwartz and Bransford 1998). Analyzing these contrasting examples help students to develop a concrete, yet varied and deep understanding of the criteria for self-assessment. In our study, such active comparisons seemed to have helped students generate the differentiated knowledge structures that enabled them to understand deeply what specific good features to include and what specific poor features to avoid in the story writing. This resulted in recognition of poor aspects of writing, generation of strategies to correct them, and willingness to revise and make improvements.

In theory, self-assessment supported by contrasting cases should benefit self-regulation. In our study, students exposed to contrasting cases were able to recognize examples of bad writing and produced better writing products in the end. However, we did not explicitly test whether the ability to produce higher quality writing in the end was a result of their conscious self-regulation due to self-assessment activity. Very few studies have examined how the process of self-assessment enhances self-regulation in writing or learning in other subjects. It is not yet known how an improved self-assessment facilitates self-regulation learning processes, and particularly at what phase of the process. It is likely that engaging in self-assessment at the planning stage may not offer as many benefits as engaging in self-assessment after the first draft is produced. Therefore, it is important to investigate how self-assessment training affects each of the three phases of self-regulation: planning, execution and revision.

Limitations of the present study

There are several limitations apparent in this work. First, this investigation was accomplished with two 6th grade classrooms in a single school and was demonstrated with a particular story-writing task. Findings are limited in scope to a particular task, age, and single instance. Additional research should replicate and extend this work in an effort to better understand how working with contrasting cases improves students' performance in other classes and to generalize findings to other contexts and multiple grade levels.

In addition, the measures of self-assessment developed for use in this study are in need of further development and validation. We did not assess if students actually revised their work as they said they would in the self-assessment. That would be a useful measure in the future. Classroom teachers often use self-assessment measures for specific purposes. What sorts of questions best capture students' self-assessment, especially for use in classroom research? Additional investigations should consider issues of validity and reliability when considering ways to measure students' self-assessment for use in classroom-based research.

A third limitation was the inability to disentangle the effects of classroom discussion from the effects of the intervention. Engaging in authentic, classroom based research creates dilemmas between researcher control and classroom authenticity. The classroom

cases were an integral part of the discussion, so we believe that the contrasting cases should be understood and considered in the context of teacher implementation and class discussion. We are not attempting to promote that using contrasting cases without appropriate teacher input is what worked here, but that the contrasting cases used by teachers in a manner natural for their discussion promoted the improvement in student work and self-assessment, more so than use of only good cases in the same discussion context did.

Future research directions

There are a number of areas ripe for further investigation. First, the fact that a simple instructional strategy, use of contrasting cases, could produce benefits suggests the need to research its usefulness in relation to other writing skills (e.g., expository, creative, opinion, etc.) as well as its effect on learning in other academic domains. Examples include: does the use of contrasting cases help students better represent and hone their understanding of complex scientific knowledge and processes? Does the use of these kinds of cases make students better at assessing their need for improvements and potentially create metacognitive awareness in domains, such as, mathematics?

A second area for future research involves instructional concerns. These issues include investigating when contrasting cases may be most helpful for students, what types of students need the contrasting case instructional support the most, and the ways teachers can best present and organize discussion around contrasting cases. Further research can better elaborate on the variations that likely exist in self-assessment for students in different grades and those with different levels of achievement.

Moreover, additional research is needed to address potential effects of contrasting cases for stimulating self-regulation and metacognition. What kinds of contrasting cases improve one's ability to notice deep features of concepts and to accurately self-assess and revise one's work? Under what conditions do contrasting cases helps students better understand abstract ideas and recognize when to use their knowledge? How transferable are they? Will students spontaneously seek out contrasting examples/cases when attempting to learn new material?

Finally, attitudinal and motivational benefits towards subject matter as the result of improved self-assessment skills should also be measured. Future studies could investigate whether effective self-assessment also improves students' self-efficacy and interest in those subjects they are asked to self-assess. Greater self-awareness in conjunction with understanding of criteria of evaluation likely helps students' attitudes and persistence. Although not investigated here, such ideas are also worthy of study.

Implications for instructional design

A key assumption underlying contrasting cases instruction is that students learn to pick up or notice important information in the environment (Garner 1974; Gibson 1969). Our methodology for designing contrasting cases involves identifying and creating relevant gradients that makes a story good by highlighting similarities and differences. Doing so will make the important features explicit to the learner. Rather than having students explore freely in the hope that they will generate the right information, contrasting cases instruction aims to provide students with a framework to discover important features for

writing and self-assessment. Furthermore, learning goals should guide decisions on what criteria or features to highlight in contrasting cases. Prioritization of multiple goals may also influence how many certain features to include or highlight in the cases. For instance, in our study, spelling correctly is not given as high priority as making the story content interesting.

Instructional designers and teachers rarely think of using contrasting cases even though there are apparent benefits for using them. A typical instructional approach involves showing multiple examples of a good model. Yet, a drawback with this approach is explained by the perceptual learning theory: learning what a thing is also depends on learning what it is not (Gibson 1969; Schwartz and Bransford 1998). Utilizing contrasting cases involves more than merely showing multiple examples multiple times, rather it involves presenting different features of the concepts and helping students to compare and contrast them in an effort to engage them in more expert-like differentiation of important features (Schwartz and Bransford 1998). Simply providing students with contrasting cases does not automatically produce deep understanding and self-assessment. Students will find these different features confusing and they will not be able to know what features are relevant and important (Schwartz et al. 2011). Students require a frame that guides their search for deep structure understanding (Lin et al. 2005; Schwartz et al. 2011) and for evaluating their work. It is important for teachers to direct students in searching for important patterns in the cases and, moreover, to guide students to relate concrete ideas in the contrasting cases to abstract criteria.

Conclusion

Perceptual learning theories originated by Gibson and Gibson (1955) have been tested in lab settings, and the findings are consistently robust, demonstrating effects on discriminating sensory material. Yet, the potential for applying perceptual learning theories to academic learning and skills in a classroom setting remain largely untapped. Using contrasting cases is an innovative instructional approach to improve writing and self-assessment of one's writing. Specifically, highlighting good and poor features of writing and using criteria to guide the discussions of these features enhance students' ability to identify specific areas for future improvement. This approach is simple and promising: Instructional materials may be created that utilize contrasting cases and can be adapted for a range of grade levels and subject areas. Doing so has the potential to improve students' understanding of academic material as well as promote self-assessment of their learning.

References

Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education, 4*(4), 31–37.

Andrade, H. G. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). New York: Routledge.

Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*(1), 21–34.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation, 10*(3), 1–11.

Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice, 17*(2), 199–214.

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*(2), 3–13.

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research, 102*(4), 287–301.

Andrade, H. L., & Warner, Z. B. (2012). Beyond "I give myself an A". *Educator's Voice, 5*, 42–51.

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56*(1), 45–72.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470–497). New York: Cambridge University Press.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77–165). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Childers, J. B. (2008). The structural alignment and comparison of events in verb acquisition. In V. S. Sloutsky, B. C. Love, & K. McRae (Eds.), *Proceedings of the 30th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Childers, J. B., & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36*(1), 201–224.

De Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*, 245–252.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in education theory and practice* (pp. 249–276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produce underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education and workplace. *Psychological Science in the Public Interest, 5*(3), 69–106.

Elder, A. D. (2010). Children's self-assessment of their school work in elementary school. *Education 3–13, 38*(1), 5–11.

Eva, K. W., & Regehr, G. (2008a). "I'll never play professional football" and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions, 28*(1), 14–19.

Eva, K. W., & Regehr, G. (2008b). Knowing when to look it up: A new conception of self-assessment ability. *Academic Medicine, 82*(10), 46–54.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395–430.

Flavell, J. H., & Wellman, H. (1977). Metamemory. In R. V. Vail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 220–241). Hillsdale: Erlbaum.

Flower, L., & Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In C. Frederikson & J. Dominic (Eds.), *Writing: Process, development and communication* (pp. 39–58). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development, 82*(4), 1173–1188.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning to transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. NY: Meredith Corporation.

Gibson, J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. *Psychological Review, 62*, 32–51.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38.

Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology, 99*(2), 297–310.

Graham, S. (2006). Writing. In P. A. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 457–478). New York: Macmillan.

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*, 3–12.

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). New York: Guilford Press.

Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*(4), 710–744.

Graham, S., Kiuhara, S., McKeown, D., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(4), 879–896.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476.

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*(3), 334–372.

Hacker, D. J., Keener, M. C., & Kircher, J. C. (2009). Writing is applied metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 154–173). New York: Taylor & Francis Group.

Harris, K. R., & Graham, S. (1992). *Helping young writers master the craft: Strategy instruction and self-regulation in the writing process*. Cambridge, MA: Brookline.

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 131–153). New York: Taylor & Francis Group.

Harris, K. R., Santangelo, T., & Graham, S. (2010). Metacognition and strategies instruction in writing. In H. S. Waters & W. Schneider (Eds.), *Metacognition, strategy use and instruction* (pp. 226–256). New York: Guilford.

Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: Does comparing with inequality symbols help? *Journal of Experimental Child Psychology, 107*(1), 15–30.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics, 55*(5), 440–454.

Higham, P. A. (2013). Regulating accuracy on university tests with the plurality option. *Learning and Instruction, 24*, 26–36.

Houghton Mifflin Social Studies Textbook Support. (1999). *A message of ancient days: Unit Activities and Resources*. Orlando, FL: Houghton Mifflin Harcourt.

Kitzantas, A., & Zimmerman, B. J. (2006). Enhancing self-regulation of practice: The influence of graphing and self-evaluative standards. *Metacognition and Learning, 1*(3), 201–212.

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction, 22*(4), 296–298.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 35–69.

Kostons, D., van Gog, T., & Paas, F. (2011). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*(2), 121–132.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning, 5*(2), 173–194.

Lin, X. D. (2001). Designing metacognitive activities. *Educational Technology Research and Development, 49*(2), 23–40.

Lin, X. D., & Bransford, J. D. (2010). Personal background knowledge influences cross-cultural understanding. *Teachers College Record, 12*(7), 1729–1757.

Lin, X. D., Schwartz, D., & Hatano, G. (2005). Toward teachers' adaptive metacognition. *Educational Psychologist, 40*(4), 245–255.

Lin, X. D., Siegler, R., & Sullivan, F. (2010). Students' goals influence their learning. In R. J. Sternburg & D. D. Preiss (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching and human development* (pp. 79–105). New York: Springer.

McCabe, A., & Peterson, C. (1984). What makes a good story? *Journal of Psycholinguistic Research, 13*(6), 457–480.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159–163.
Moskal, B. M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research and Evaluation*, *8*(14). Retrieved April 12, 2012 from http://PAREonline.net/getvn.asp?v=8&n=14.
Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.
Norcini, J. (2003). Peer assessment of competence. *Medical Education, 37*(6), 539–543.
Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*, 92–104.
Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407.
Pintrich, P. R., Conley, A. M., & Kempler, T. M. (2003). Current issues in achievement goal theory and research. *International Journal of Educational Research, 39*(34–5), 319–337.
Pressley, M., & Harris, K. R. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 265–286). New York: Macmillan.
Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology, 35*, 28–43.
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574.
Rittle-Johnson, B., & Star, J. R. (2009). Compared to what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544.
Sargeant, J. (2008). Toward a common understanding of self-assessment. *Journal of Continuing Education in the Health Professions, 28*(1), 1–4.
Sargeant, J., Mann, K., van der Vleuten, C., & Metsemakers, J. (2008). "Directed" self-assessment: Practice and feedback within a social context. *Journal of Continuing Education in the Health Professions, 28*(1), 47–54.
Scardamalia, B., & Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Current research and open questions* (pp. 563–577). Hillsdale, NJ: Lawrence Erlbaum Associates.
Schneider, P., & Winship, S. (2002). Adults' judgments of fictional story quality. *Journal of Speech, Language and Hearing Research, 45*(2), 372–383.
Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*, 33–45.
Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475–522.
Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759–775.
Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.
Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024–1037.
Thompson, L., Gentner, D., & Loewenstein, J. (2000). Avoiding missed opportunities in managerial life: Analogical training more powerful than individual case training. *Organization Behavior and Human Decision Processes, 82*, 60–75.
Tsivitanidou, O. E., Zacharia, Z. C., & Hovardas, T. (2011). Investigating secondary school students' unmediated peer assessment skills. *Learning and Instruction, 21*(4), 506–519.
Van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merrienboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction, 24*, 15–25.
VanLehn, K., & Van De Sande, B. (2009). Acquiring conceptual expertise from modeling: The case of elementary physics. In K. A. Ericsson (Ed.), *The development of professional performance: Toward measurement of expert performance and design of optimal learning environment* (pp. 356–378). Cambridge, UK: Cambridge University Press.

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 1197–1218). New York: Routledge.
Wang, S.-H., & Baillargeon, R. (2008). Can infants be ''taught" to attend to a new physical variable in an event category? The case of height in covering events. *Cognitive Psychology, 56*(4), 284–326.
Winne, P. H. (2005). Key issues in modeling and applying research on self-regulated learning. *Applied Psychology: An International Review, 54*(2), 232–238.
Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology, 61*, 653–678.
Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego: Academic Press.
Zimmerman, B. J. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 705–722). Cambridge, MA: Cambridge University Press.
Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*, 166–183.
Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660–668.
Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22*(1), 73–101.
Zimmerman, B., & Schunk, D. H. (2011). *Handbook of self-regulation of learning and performance*. New York: Routledge.

Xiaodong Lin is interested in motivation, instructional design and science learning

David Shaenfield is interested in argumentation, science education and instructional design

Anastasia D. Elder is interested in self-assessment, instructional design, science learning, and middle level students