

Examining response confidence in multiple text tasks

Alexandra List¹ · Patricia A. Alexander¹

Received: 6 February 2014 / Accepted: 7 April 2015 / Published online: 6 May 2015
© Springer Science+Business Media New York 2015

Abstract Students' confidence in their responses to a multiple text-processing task and their justifications for those confidence ratings were investigated. Specifically, 215 undergraduates responded to two academic questions, differing by type (i.e., discrete and open-ended) and by domain (i.e., developmental psychology and astrophysics), using a digital library of seven varied texts (e.g., journal article, newspaper, Wikipedia). Students then indicated how confident they were in their responses and provided justifications for their level of confidence. This investigation had four main goals: (a) to determine the strength of students' response confidence; (b) to understand the criteria students use to derive their confidence ratings; (c) to consider the effect of task conditions on ratings and justifications; and (d) to examine the relation between students' response confidence ratings and justifications and the accuracy and quality of their responses. Students' justifications for response confidence were found to vary across task conditions. Further justifications for response confidence based on sources consulted (i.e., text-directed justifications) were found to relate to response accuracy for the discrete question and to response quality for the open-ended question.

Keywords Response confidence · Justifications · Task type · Multiple source use · Calibration

The process of obtaining and integrating information from multiple documents to meet task demands, referred to as multiple source use (MSU), is ubiquitous in students' academic lives (Goldman and Scardamalia 2013; Purcell et al. 2012). Despite its prevalence, multiple source use is not without its challenges (Goldman et al. 2012; Strømsø et al. 2008), even for those who have been characterized as *digital natives* (Prensky 2001). For example, Rouet (2006) describes the multiple source use process as presenting greater cognitive and informational demands for students, as compared to interactions with a single text. Further, Bråten (2008) refers to the challenges of meeting task demands using multiple texts as posing a complex, or high-level, problem for learners.

✉ Alexandra List
alist@umd.edu

Patricia A. Alexander
palexand@umd.edu

¹ University of Maryland, College Park, MD 20742, USA

Indeed, students, even at the undergraduate level, have been found to experience difficulties with multiple source use (Currie et al. 2010; Grimes and Boening 2001; Metzger et al. 2003). Moreover, these difficulties appear to occur throughout the MSU process, including during text selection (Salmerón and Kammerer 2012; Walraven et al. 2009) and evaluation (Bråten et al. 2011; Mason et al. 2010), and to extend to response composition (Bråten 2008; Grimes and Boening 2001; Stahl et al. 1996).

Despite documented difficulties, few researchers have turned to students themselves to examine whether they are aware of these challenges and their origins. For instance, are students confident that they have responded well to particular questions and how do they make such a determination? To our knowledge, no researchers have explicitly investigated the level of students' confidence in their responses to tasks requiring the use of multiple texts. Nor are we aware of studies that have directly considered students' judgments about the accuracy or quality of their responses to MSU tasks vis-à-vis the actual accuracy and quality of those responses, or students' *calibration* (Glenberg and Epstein 1987) when completing MSU tasks.

Indeed, the criteria students use to determine ratings of response confidence have only begun to be explored in studies of single text comprehension (Dinsmore and Parkinson 2013; Pressley et al. 1990). Thus, the purpose of this study was to examine students' level of response confidence when completing multiple source use tasks, to probe the criteria they apply when reaching judgment of confidence, and to juxtapose these perceptions against the reality of students' written performance. We were also interested in the extent to which students' response confidence might differ depending on the nature of task conditions; specifically, the type of questions posed and the domains queried.

Defining response confidence

Response confidence, alternately referred to as response certitude or response certainty (Vasilyeva et al. 2008a, b), represents students' *subjective certainty* in the accuracy of their selected responses or constructed answers (Kulhavy and Stock 1989; Kulhavy et al. 1990). In determining response confidence, students are thought to engage in a process of comparing their own answers to some well-defined ideal, often conceptualized as a single "correct" answer that can be objectively scored (Kulhavy and Stock 1989).

Yet, today's students are increasingly asked to complete academic tasks for which such an ideal may not exist (Goldman and Scardamalia 2013; Hartley and Bendixen 2001). Indeed, students are often tasked with finding and integrating information across multiple texts for the purpose of crafting responses to more open-ended questions, judged on the basis of rather complex and subjective criteria (Goldman 2004). There is a need to understand not only how students determine their level of response certitude when answering discrete questions (i.e., those having a single agreed-upon answer), but also how they make appraisals of more open-ended, elaborated responses (i.e., responses for which the standards of what constitutes an accurate, quality answer are less evident).

Temporality and response confidence ratings

In studies of response certitude, participants have been asked to rate their confidence either prior to engaging with a task or following task completion (Pieschl 2009). When students

make a priori judgments of whether they will perform successfully on a task or assessment, these evaluations have been termed *judgments of confidence* (Hadwin and Webster 2013; Schraw 2009) or *judgments of learning* (Nelson and Dunlosky 1991; Vidal-Abarca et al. 2010). Much of the current research has focused on these types of a priori judgments (Hadwin and Webster 2013; Koriat 2012).

In the present study, our focus was on students' post-hoc judgments of confidence; that is, ratings of response certainty following task completion. We elected to focus on students' post-hoc judgments, as we were interested in students' self-evaluations once a task had been completed. Such post-hoc judgments are part of the multiple source use process outlined in the MD-TRACE model (Rouet and Britt 2011, see next section). In addition, studies of calibration have suggested that students' post-hoc judgments, termed *post-dictions*, may be more stable and accurate than a priori predictions of performance (Hadwin and Webster 2013; Maki and Serra 1992; Pieschl 2009).

Response confidence in MSU tasks

We can look to models of MSU to offer guidance in understanding how students may make judgments of confidence in their responses to more open-ended, ill-structured questions. One such prominent model is the Multiple-Documents Task-Based Relevance Assessment and Content Extraction Model (MD-TRACE; Rouet and Britt 2011), which conceptualizes multiple source use as unfolding in a series of five steps. In Step 1, students build a *task model*, a cognitive model of perceived task demands and a plan for how these demands may be satisfied. In subsequent steps, students determine a need for information, thereby electing to consult texts (Step 2); interact with these texts (Step 3); and, develop a written response (Step 4). The ultimate step of this model (Step 5) involves students examining their written responses to determine whether they satisfy task demands. Ratings of response confidence, within an MD-TRACE framework, occur at this fifth and final step. In determining the correspondence between their own written products and some perceived ideal, students formulate beliefs about the accuracy and quality of their responses (i.e., response confidence).

It may be more precise to say that in Step 5 students are comparing their written response to their *perception* of task demands, or to the task model initially constructed in Step 1 (Rouet and Britt 2011). Students' task model and subsequent judgments of response quality in Step 5 of the MD-TRACE model may be based not only on understandings of question demands but also on conceptions of what level of response may be sufficient to meet task demands and on what contextual limitations may be at play (e.g., limitations in time or effort). Thus Step 5 represents a negotiation between students' conceptions of an optimal response and pragmatic concerns or limitations.

According to the MD-TRACE model, the comparisons students make between their written response and their understanding of task demands in Step 5 are not binary (i.e., not match versus no match determinations). Rather, the comparison between constructed response and task model in Step 5 can be characterized as students' relative certitude that indeed task goals have been met. Based on insights from the MD-TRACE model, we contend that response confidence ratings have three primary features. Specifically, response confidence ratings appear to be:

1. Subjective and individually determined;
2. Task specific, reflecting the correspondence between perceived task demands and students' given responses; and,
3. Varied in extent (i.e., strength of the belief) and in origin (i.e., basis for the belief).

Origins of response confidence ratings

Students are thought to determine their level of response confidence based on two sources of information (Kulhavy and Stock 1989). First, students consider *task criteria*, or information specific to and provided within the context of a particular task. This includes determinations of confidence based on task demands or on information contained in task-related texts. Second, in judging their response confidence, students continuously make comparisons between their own responses and *cognitive referents*, or information external to the specific task, that nonetheless comes to bear on response evaluation (Kulhavy and Stock 1989). Cognitive referents include students' prior knowledge of and experiences with similar texts and tasks. Thus, even when task demands do not provide explicit criteria for response evaluation, students can nonetheless draw on their cognitive referents, or prior experiences completing similar tasks, to derive potential standards for assessing response quality. For example, while an open-ended task, such as researching the effects of climate change, may not explicitly ask students to consider text reliability, based on prior experiences completing similar tasks, students may nonetheless know to select authoritative sources of information.

While these two dimensions (i.e., task criteria and cognitive referents) have been identified as the origins of students' response confidence ratings, in the literature, students' confidence ratings have not been differentiated with regard to these dimensions. More generally, in asking students to rate response confidence, the origin of or bases for these ratings have been under-examined. At the same time, Dinsmore and Parkinson (2013) identified the importance of considering not only the absolute value of students' confidence judgments but their content as well, assessed via confidence judgment explanations or justifications. Insights into the bases of students' confidence judgments may be particularly important in understanding the nature of limitations in students' calibration. Examining single text tasks, Dinsmore and Parkinson (2013) classified students' justifications for response confidence into four categories: prior knowledge, characteristics of the text, characteristics of the item, and guessing, with students most frequently justifying response confidence based on text characteristics or offering text-directed justifications.

To add further nuance to the confidence justification categories identified by Dinsmore and Parkinson (2013), we were interested in examining the extent to which these reflected either epistemic or non-epistemic determinations. In effect, given their inherent complexity (Bråten 2008; Rouet 2006), multiple text tasks have generally been viewed as contexts within which students' epistemic beliefs may manifest and be particularly determinant of behavior (e.g., Mason et al. 2010, 2011). Most commonly, students' epistemic beliefs have been thought to entail judgments of texts or text evaluations (e.g., Bråten et al. 2008). However, in the present study we were interested in examining the extent to which students' judgments of response confidence were likewise impacted by epistemic determinations, particularly given work identifying the role of epistemic beliefs in self-regulated learning (e.g., Greene et al. 2010; Muis 2007). For instance, we were interested in the extent to which students' judgments of text

reliability (i.e., judgments related to students' beliefs about source of knowledge, Mason et al. 2011; Mason et al. 2010) were cited as text-directed justifications for response confidence.

Task and response confidence

In general, studies of response confidence ratings have assumed students to be operating in a multiple-choice framework (Kulhavy et al. 1990; Moore 2007; Vasilyeva et al. 2008b). Thus, in determining their response confidence students have been thought to be making appraisals of the accuracy of a limited number of potential answer choices provided to them. Kulhavy et al. (1979) refer to this process of ranking the potential correctness of a restricted number of answer choices as constructing a *hierarchy of confidence*. However, in today's schools, students are increasingly expected to complete more complex tasks, generate their own answers, and evaluate a large numbers of highly varied potential responses. Typical of multiple source use tasks, students are asked to produce elaborated responses, evaluated on multiple, implicit criteria (Rouet and Britt 2011).

For these more open-ended questions, often requiring the use of multiple information sources, confidence judgments would seem to be more difficult for students to determine. First, the inclusion of multiple texts introduces a larger body of information that students must consider in developing their judgments of response confidence. Further, given that open-ended tasks are characterized as ill defined in terms of both what is asked of students and which evaluative criteria are applied (Alexander 2006), students may need to rely to a greater extent on prior task experience to determine the criteria upon which to base their response evaluations. Mosenthal (1998) suggests that the complexity associated with open-ended tasks stems from the plurality of potential answers that may be generated by making varied connections across texts and from the varied and flexible standards applicable to judging a quality response. In turn, these features may complicate students' judgments of response confidence (Pieschl 2009; Winne and Hadwin 1998).

Studies of confidence judgments have been critiqued for neither sufficiently considering students' subjective perceptions of task demands, termed *frames of reference*, nor fully examining students' awareness of the external criteria along which their responses may be judged (Pieschl 2009), suggesting that examining students' response confidence across task conditions may be of particular import. Further, to fully understand students' perceptions of task demands requires examining not only their confidence ratings but also the justifications for these ratings.

Domain and response confidence

In addition to considering variations in students' response confidence ratings and justifications across question types, we were also interested in examining response confidence ratings across domains, namely, developmental psychology and astrophysics. In studies of calibration, higher performing students demonstrate better calibration (Grimes 2002; Hadwin and Webster 2013), as do students taking advanced, rather than introductory, courses (Falchikov and Boud 1989; Grimes 2002). The more accurate calibration demonstrated by more advanced students may stem from either their greater knowledge of a domain, and therefore greater knowledge of the criteria along which responses may be evaluated, *or* from greater knowledge of themselves as

learners, and therefore of their prior performance in a given discipline. In either case, the role of prior domain experience in students' confidence judgments is important to consider (Maki and Serra 1992) particularly as the multiple source use process has been characterized as task-driven (Rouet 2006; Rouet and Britt 2011).

Present study

The present study sought to contribute to the literature by examining the strength and accuracy of students' response confidence ratings when responding to questions that required the use of multiple digital texts and students' justifications for these ratings. Undergraduate students were asked to respond to two types of questions: discrete, requiring a single, concrete answer, and open-ended, requiring a more elaborated response and allowing for greater variation in quality, across two domains (i.e., developmental psychology and astrophysics). After answering each question, students were asked to rate and justify their response confidence. The particular questions we set out to address in this research were the following:

1. What kinds of justifications do students offer for their response confidence ratings when responding to discrete and open-ended questions in developmental psychology and astrophysics?

Similar to the Dinsmore and Parkinson (2013) study, it was expected that students would offer justifications for confidence judgments related to information provided in texts, their prior knowledge, and aspects of the task. Further, as students were placed in a multiple text context and instructed to make use of the library, it was expected that the majority of their response confidence judgments would be text-directed. Consistent with this hypothesis, Dinsmore and Parkinson (2013) found the majority of students' confidence justifications, in a single text context, to be text-directed. In the current study, we expected this effect to be even more pronounced when students were asked to justify their response confidence in a multiple text task.

2. To what extent do students' response confidence ratings and justifications offered differ across question type (i.e., discrete versus open-ended) and academic domain (i.e., developmental psychology versus astrophysics)?

In examining ratings of response confidence and their justifications across task conditions, a main effect for both question type and domain was expected. Based on prior research, we hypothesized that students would express more confidence in responding to the discrete questions, with a single specific answer, than to more open-ended questions, for which standards of response quality may have been more ambiguous and therefore more difficult to determine (Mosenthal 1998). Additionally, due to our participants' greater experience with developmental psychology, rather than astrophysics, it was expected that they would have great confidence in the more familiar domain. Thus, we expected students to report the greatest confidence in the discrete question in developmental psychology and to express the least confidence in open-ended astrophysics questions.

Further, we expected students to produce more text-directed justifications for response confidence for their answers to discrete questions. It was expected that students would have perceived the discrete questions as having a single correct answer that could be

identified in the texts. Additionally, given participants' social science backgrounds, it was expected that students would produce more personally directed epistemic justifications in response to developmental psychology questions. We expected students to feel as if they "knew" the answers in developmental psychology. On the other hand, as students had more limited knowledge in the domain of astrophysics, they were expected to justify their confidence ratings in that domain based on information provided in texts rather than based on prior knowledge.

3. In what ways do students' ratings of response confidence and the justifications they provide for those ratings relate (a) to the accuracy of their responses to discrete questions, and (b) to the characteristics of their responses to open-ended questions in developmental psychology and astrophysics?

Based on the literature suggesting that students are poorly calibrated (e.g., Dunlosky and Rawson 2012; Maki et al. 2005), we expected the associations between response confidence and response performance to be somewhat limited. With regard to response justifications, we hypothesized students responding correctly to the discrete question and with higher levels of response quality, in the case of the open-ended question, to produce more text-directed justifications. It was generally expected that consulting texts and justifying response confidence based on text quality would be associated with better responses. Particularly for the open-ended question, it was hypothesized that students' citing task-directed justifications would have stronger responses, as these types of justifications seem indicative of greater sensitivity and concern for task demands.

Methods

Participants

Participants in this study were 215 undergraduate students at a large Mid-Atlantic university in the United States. Ten of these participants were ultimately excluded because they did not report a rating of response confidence for either of the two questions. Of the final 205 participants, 69.76 % were female ($n=143$; male: 29.76 %, $n=61$) with a mean age of 20.70 ($SD=2.62$). The students were majority White (63.41 %, $n=130$), with 13.17 % of students reporting Asian ethnicity ($n=27$), 12.20 % reporting African American ethnicity ($n=25$), and 4.88 % ($n=10$) self-identifying as Hispanic/Latino. In addition, 3.90 % of students categorized themselves as multi-racial ($n=8$) and four students declined to report their ethnicity. Further, 8.78 % of students were non-Native English speakers ($n=18$). One participant did not provide demographic information. Students received extra credit for participation.

The undergraduate students in this study represented a variety of majors, particularly in the social sciences (73.66 %, $n=151$), including psychology, economics, and hearing and speech. Additionally, 21.46 % of students majored in the natural sciences ($n=44$) and 3.90 % of students ($n=8$) majored in the humanities. The participating students included freshmen (1.95 %, $n=4$), sophomores (24.88 %, $n=51$), juniors (31.22 %, $n=64$), and seniors ($n=37.07$ %, $n=76$) with 4.39 % ($n=9$) of students enrolled for post-baccalaureate credit. The average self-reported GPA was 3.21 ($SD=0.43$) on a 4.0 scale.

Materials

For this study, participants were provided with a library of seven digital texts, specific to each of four questions assigned. The texts included in each of four libraries varied in reliability and source type and included a journal article, a government report, a news site, featuring popular press summaries of research studies, a newspaper article from the *New York Times*, a corporate-sponsored website, a Wikipedia page, and a blog. These texts represented the types of information sources students might encounter while conducting research on the Internet. As seen in the screenshot displayed in Fig. 1, each text was reformatted to appear uniform. At the top of each text, source information was provided, including publication source, logo, and title, with text below. In reformatting, all images were removed and hyperlinks disabled.

All texts in a particular library were relevant to the target question (i.e., containing information potentially useable in crafting a response), but varied in quality. The texts also offered different perspectives on the target issue (Lenski 1998)—perspectives that could be complementary or conflicting. For example, when asked to research factors promoting linguistic development in young children, students' responses could have been informed by biological and physiological perspectives on development (government report; blog), a report on the role of environmental factors, such as noise in the home, on vocabulary acquisition (news site), or a study exploring sociological factors, like race and socioeconomic status, in children's early language use (journal article). Each of the texts contained a potential answer to the question. Although questions could have been answered with information from only a

Search Task Question 1 of 2

Please answer as you would for an academic class using the Source Library.

Which factors may promote or hinder verbal development in young children? Please explain.

The Role of Home Literacy Practices in Preschool Children's Language and Emergent Literacy Skills

The Role of Home Literacy Practices in Preschool Children's Language and Emergent Literacy Skills

Next Question

Source Library

Government Report Journal Article Newspaper News Site Online Encyclopedia Blog Website



AMERICAN
SPEECH-LANGUAGE-
HEARING
ASSOCIATION

JSLHR

Journal of Speech, Language, and Hearing Research

The Role of Home Literacy Practices in Preschool Children's Language and Emergent Literacy Skills

This study examined how 4 specific measures of home literacy practices (i.e., shared book reading frequency, maternal book reading strategies, child's enjoyment of reading, and maternal sensitivity) and a global measure of the quality and responsiveness of the home environment during the preschool years predicted children's language and emergent literacy skills between the ages of 3 and 5 years. Study participants were 72 African American children and their mothers or primary guardians primarily from low-income families whose home literacy environment and development have been followed since infancy. Annually, between 18 months and 5 years of age, the children's mothers were interviewed about the frequency they read to their child and how much their child enjoyed being read to, and the overall quality and responsiveness of the home environment were observed. Mothers also were observed reading to their child once a year at 2, 3, and 4 years of age, and maternal sensitivity and types of maternal book reading strategies were coded. Children's receptive and expressive language and vocabulary were assessed annually between 3 years of age and kindergarten entry, and emergent literacy skills were assessed at 4 years and kindergarten entry. The specific home literacy practices showed moderate to large correlations with each other, and only a few significant associations with the language and literacy outcomes, after controlling for maternal education, maternal reading skills, and the child's gender. The global measure of overall responsiveness and support of the home environment was the strongest predictor of children's language and early literacy skills and contributed over and above the specific literacy practice measures in predicting children's early language and literacy development.

Understanding how the home literacy environment affects the acquisition of children's later language and emergent literacy knowledge has become of increasing interest during the past three decades (Lonigan, Burgess, & Anthony, 2000; Neuman & Dickinson, 2001; Senechal & LeFevre, 2002; Whitehurst & Lonigan, 1998). Children who acquire successful initial reading skills tend to remain good readers, while children who experience difficulty in learning to read tend to continue to have problems in reading throughout the school years (Adams, 1990; Baydar, Brooks-Gunn, & Furstenberg, 1993; Cunningham & Stanovich, 1997; National Research Council, 1998). Considerable research has shown children's emergent literacy skills—the skills, knowledge,

Fig. 1 Screenshot of search task interface

single text, multiple texts were required to corroborate answers to the discrete question or to craft comprehensive and integrative responses to the open-ended question.

Texts were naturally occurring and identified by searching Google for both the exact wording of the target question and key words within the target question (e.g., verbal development). As such, texts varied with regard to length and readability but were considered to be appropriate for use with an undergraduate audience. Texts were then selected based on target criteria including source type, publisher or origin (e.g., New York Times), and match across the four question types. Readability and word count statistics are provided in the [Appendix](#). Texts were piloted with undergraduate students to ensure readability and usability in responding to target questions.

Students were free to access none, some, or all of the texts in any order and they could keep texts open while responding to each question. To encourage students' deliberative text selection, texts could be accessed through a link listing only the source type. For example the "newspaper" link was used to access the New York Times article. Links were arranged in alphabetical order by source type. In responding to the study questions, students were instructed to answer as they would for an academic class. There was no time limit placed on the task. However, students had to complete answering the first question before being presented with the second.

Measures

Participants were asked to respond to two questions using the described library of digital texts specific for each question. In particular, they were asked to respond to one question that was discrete and one question that was open-ended. One of those questions came from the domain of developmental psychology and the other from astrophysics. For example, one group of participants may have received the questions, "How many words are in a normally developing 24-month-old's vocabulary," (discrete, developmental psychology) and, "Which planetary features may promote or hinder habitability? Please explain," (open-ended, astrophysics) in counterbalanced order. Alternately, participants could have received the questions, "How many stars are in the Milky Way galaxy," (discrete, astrophysics) and, "Which factors may promote or hinder linguistic development in young children? Please explain," (open-ended, developmental psychology) counterbalanced. As such, both question type (i.e., discrete and open-ended) and domain (i.e., developmental psychology and astrophysics) were counterbalanced in the tasks presented to students. In the case of both the discrete and open-ended question participants were required to enter a response. However, while in the case of the discrete question, the response involved a singular answer, in the open-ended question, a more elaborative response was expected.

These questions were intended to represent the types of queries students may be assigned or elect to research during the course of an introductory class in each respective domain. The domains of developmental psychology and astrophysics were selected because they represented disciplinary differences associated with the social and natural sciences. Further, given that our sample was primarily social science majors, with some representation of natural science fields, we expected students' knowledge and interest in the chosen domains and the specific topics of inquiry to vary and to differentially impact determinations of response confidence.

Response coding Students' responses to the discrete questions were coded as a binary, either correct or incorrect, based on information provided in the texts (i.e., astrophysics: 200–400

billion starts; developmental psychology: 50–250 words in a 24-month olds vocabulary). Responses were coded as correct if students either reported the correct range or the upper or lower bound reported in texts (e.g., “at least 50”). In the case of the open-ended questions, five indices to assess response quality were used. These indices were informed by prior investigations of MSU (e.g., Gil et al. 2010). First, the number of words included in students’ open-ended responses was counted. Although word count is not a measure of response quality per se, it was nonetheless considered as a characteristic of students’ responses as response length has been examined in prior research on MSU (e.g., McNamara et al. 2010). Further, students in our study provided response confidence justifications based on amount written, suggesting that response length may be an important indicator for students. Next, we counted the number of pieces of *evidence* students introduced in their responses. For this determination, a piece of evidence was defined as a unique or distinct reason offered to substantiate a response. In addition to considering the number of pieces of evidence, or reasons, students provided, we were also interested in the quality of these reasons. Thus, we also counted the number of *elaborations* students included in their responses. Elaborations were pieces of evidence that contained additional information that extended beyond a given claim and justification. Such elaborations included additional explanations, examples, inferences, or generalizations. For example, one students’ response included the statement that stimuli contribute to verbal development: “Stimuli such as talking toys, human interaction, exposure, and television because they promote communication between individuals.” While stimuli promoting verbal development by fostering communication was coded as students providing a single piece of evidence, the examples provided also meant this evidence was elaborated.

In addition, students’ responses were coded according to a modified SOLO Taxonomy (Biggs and Collis 1982). The SOLO taxonomy is a response rating scale, ranging from zero to four. As in prior studies (e.g., Biggs 1979), we modified the scale slightly to include half point increments or *transitional* levels to better reflect variations in students’ responses and to increase the information this scale provided. The SOLO Taxonomy considers not only the amount of information included in responses and their degree of integration, but also the internal and external consistency of responses when rendering a score. A score of one indicated that students included a single piece of relevant information in their response, whereas a score of two indicated that students’ answers included multiple pieces of information to substantiate responses, albeit with no connection between those pieces. To receive a score of three, students’ responses had to provided multiple pieces of relevant information that were expressly connected or integrated. Finally, a score of four on the SOLO taxonomy was for responses that provided multiple pieces of information that were both integrated and externally consistent, meaning these answers considered potentially conflicting evidence or weighed the relative merits of conflicting claims.

In our modification of the SOLO Taxonomy, half points were assigned if students were aiming for a certain level of response quality, but were not entirely successful in reaching this desired level. For example, if students were attempting to provide multiple pieces of relevant information for a given question, but only one of their reasons was fully articulated, their response would receive a score of 1.5. Taken together, these assessments of students’ open-ended responses considered both information quality (i.e., the volume of evidence presented and the quality of this evidence) and response quality (i.e., extent to which evidence was elaborated and integrated). Inter-rater reliability for coding open-ended responses was determined based on a scoring of 35 participants’ responses (17.07 % of the sample). In identifying the number of pieces of evidence, Cronbach’s $\alpha=0.75$, while for identifying the number of

pieces of evidence elaborated, Cronbach's $\alpha=0.92$. Finally, for scores on the SOLO taxonomy, Cohen's kappa was $\kappa=0.68$, corresponding to *substantial* agreement (Fleiss 1971; Landis and Koch 1977; Viera and Garrett 2005).

The number of citations included in students' responses was recorded. Citations were considered to be any specific reference students made to the title, author, or source of any text provided in the library or any references made to second-order sources, such as studies discussed within texts. Students were not explicitly instructed to include citations in their responses, so the number of citations recorded represented spontaneous sourcing. Table 1 offers an overview of the scoring procedure with corresponding samples of student responses.

Response confidence After responding to each question, students were asked to rate their response confidence on a 100-mm line. Specifically, participants were asked to respond to the question, "How confident are you in your response," using a scale ranging from *not at all* to *very much so*. Next, participants were instructed to "Please explain;" that is, they were asked to justify their confidence. The screenshot of what students encountered is displayed in Fig. 2.

Justifications for ratings of response confidence were coded in three phases. In the first phase, students' justifications were segmented into utterances, with each utterance corresponding to a single, distinct thought or idea. Resultant segments generally corresponded to statements or sentences with a single verb. However, statements connected via conjunctions (e.g., *however*) were typically coded as two separate utterances. For example, the justification: "I am confident in my response because it was taken directly from the article I read, however, the other articles may have said different things," was coded as including two utterances. Next, each utterance was coded as referencing one of three possible targets: *text-directed*, *task-directed*, or *personally-directed*. These primary categories were based on Dinsmore and Parkinson's (2013) coding of students' confidence justifications.

While Dinsmore and Parkinson examined confidence justifications rendered in reference to performance on multiple-choice items completed following single text use, the present coding scheme was adjusted to reflect a multiple text context. Utterances coded as text-directed were justifications for response confidence that were based on the texts in the digital library or the information within those texts. For instance, a response such as, "I believe the source that gave me the information was very reliable," would have been coded as a text-directed justification. Those utterances coded as task-directed were justifications for response confidence that referenced either the target question per se, (i.e., the task demands) or students' generated responses (i.e., students' responses to meet task demands). For example, the response, "More confident because I only gave estimates with no exact answer," would have been coded as a task-directed because it offered a justification based on the student's perception of the task (i.e., that it did not demand an exact answer). Confidence justifications were coded as personally-directed when they referenced factors extrinsic to the materials presented in the study or elements of the given question or task, as when a respondent relied expressly on her prior knowledge of the domain in formulating her justification (e.g., "Watched accurate scientific TV shows on our Universe.")

As a general convention, when students' justifications appeared to reference aspects of both text and task, text-directed justification was the default coding. For example, a justification of limited response confidence, in which a participant stated: "I didn't use all of the sources to verify my answer," was coded as text-directed. In this case, the student was considering both the number of texts used for corroboration (i.e., text-related) and conceptualizing their response as requiring verification (i.e., task-related). However, a text-directed coding was

Table 1 Sample open-ended response coding

Sample response	SOLO score	Explanation	Pieces of evidence	Elaborations
<i>A factor that could promote verbal development is the parent's profession. If a child has a professional parent their vocabulary would have more words than a child on welfare (New York Times).</i>	1	Single relevant piece of information	1	1
SES, typical development, culture	1.5	Attempt to provide multiple reasons, however these are not fully articulated.		0
The way that adults surrounding the child communicate with him or her may hinder their development as well as the type of television shows they are exposed to.	2	Multiple relevant pieces of information, not integrated	2	0
Research shows the strongest predictor of verbal development is overall responsiveness and support of the home environment. <i>Verbal development can be promoted when children frequently are read to, although the results are modest. Another important factor is seen to be children's interest in the reading activity, although research is limited in this capacity.</i> In addition, mother's communicative strategies while reading have also been seen to influence verbal development.	2.5	Multiple relevant pieces of information with attempted integration, but reasons are listed.	4	2
There are several factors that can accelerate language acquisition in children, such as using infant directed speech or reading to the child. <i>If a parent does not interact with the child and speak to the child, such as in a maladaptive relationship, the child can miss his sensitive period for verbal development; the case study of Genie is an example of a child missing her sensitive period for normal language development.</i>	3	Multiple relevant pieces of information, integrated, and internally consistent.	3	1
Parental influences are very important in promoting proper verbal development in young children. The articles talks about the importance of brain development and nutrition, and both of these affect child verbal development. <i>However, these are both able to be managed by the parental figures in the child's life. Parenting is the most important kind of influence to promote verbal development.</i> The importance of reading and proper education is important as well, <i>and parents can help promote proper and early reading. Personally this is what I experienced by homeschooling, and parenting is essential in childhood growth and development.</i> <i>The home environment is the most important factor in verbal development because young children will learn and have the strongest ability for mental growth in early childhood, and by having an opportunity and good parenting teaching by the parents they will perform better.</i>	4	Multiple relevant pieces of information, integrated, internally consistent, and externally consistent (i.e., holds even with presentation of conflicting evidence)	6	2

Note: Pieces of evidence that are elaborated are italicized

Search Task Response Evaluation

How confident are you in your response?

Not at all | X | Very

Please explain.

I'm pretty confident in my response because of the information given to me from the source library all stated similar answers. Also from my prior knowledge on the subject matter also agreed to the information provided from the sources and allowed me to believe the information given from the sources were correct or truthful.

Fig. 2 Screen shot of response confidence rating interface

applied, as this justification specifically indicated that verification ought occur through text use. As with this example, when text use was specifically mentioned in confidence justifications, this was privileged in coding.

In the final coding phase, justifications within each reference category were classified as being either epistemic or non-epistemic in nature. Epistemic justifications focused on evaluations of the quality, accuracy, or trustworthiness of students' responses or the content upon which those responses were based. For example, one student stated that: "the source was credible." Because of this evaluation of text credibility, this justification was coded as epistemic. Conversely, non-epistemic justifications focused on more surface or superficial features of students' responses or the information those responses were based on, such as their length. When one student justified her response confidence by stating that, "article was very easy to read and find," the statement was classified as non-epistemic.

More specifically, when students provided *epistemic text-directed* justifications their response confidence ratings were based on evaluations of the trustworthiness or authority of texts in the digital library or the information within them. For example, one student justified her response confidence as, "I have always been told journal articles are the most trustworthy sources, since they are primary sources. However I only read one journal article, so I could not compare their data to anyone else's," indicating a rating of response confidence based on two types of justifications: an evaluation of text trustworthiness as well as a desire for corroboration. Utterances coded as *non-epistemic text-directed* provided justifications for response confidence based on non-epistemic or superficial features of texts, such as their length or readability. Justifications coded into this category also included confidence ratings based on finding an "exact" answer precisely stated in the texts.

Task-directed epistemic justifications were those that evaluated response quality in reference to question demands or to the level of substantiation or evidentiary quality that the task demanded. Also included in this category were justifications for response confidence that considered the accuracy, quality, or substantiation of students' written responses or demonstrated a personal conviction or belief in the correctness of an answer. For example, while one student justified her response confidence with reference to the nature of the question, saying, "this was more factual," another student justified her response confidence by focusing on the written product, explaining, "I believe what I have written is accurate and represents

empirically supported views.” *Task-directed non-epistemic* justifications were those based on superficial aspects of task or response, absent a concern for quality or accuracy of the responses produced. Included in this coding were students’ response confidence ratings based on considerations of contextual or situational aspects of task completion, such as time constraints.

Participants’ justifications for responses confidence were coded as *personally-directed epistemic* when they relied on prior knowledge, experience, or personal expertise to substantiate response accuracy. Included in this coding were responses from students who justified their confidence based on their understanding of popular consensus or reliance on an external authority. Justifications in this category were based on the expertise or external validation that students brought to the task, rather than on information accessible as a part of the activity. For instance, one student justified his response confidence by relying on his prior experience in a domain: “I am currently in a family science class, as well as a human development course,” while another based her response confidence on personal experience, “I am relatively confident in my response because...I also have a younger brother who I interact with greatly.”

Those justifications that were coded as *personally-directed non-epistemic* similarly concerned students’ prior experiences. However, in this instance, students did not describe these experiences as leading to a knowledge-based personal justification for response certainty. Rather, justifications in this category discussed students’ familiarity or interest in the task or domain as the reasoning behind their ratings of response confidence. For example, one participant justified her confidence based on her experience performing similar tasks: “I am familiar with this topic and the style of research done in the social sciences. Because of my familiarity and interest I knew exactly where to look for an answer.” In this category, we also included students’ justifications based on motivational factors, such as interest, as when one student responded, “it was much easier to find information because I am much more interested in the topic.” While factors such as disciplinarily familiarity and topic interest may be considered to be associated with students’ knowledge, these were coded as non-epistemic, unless students’ knowledge as a response confidence justification was explicitly cited. Additional samples of all the justification types by epistemic classification can be found in Table 2. Two raters coded 47 participants’ (22.93 % of the sample) justifications for response confidence. Across these 47 participants, a total of 118 justifications were coded to establish interrater agreement. In classifying students’ justifications as text-, task-, or personally-directed Cohen’s kappa was $\kappa=0.69$. In classifying students’ justifications as epistemic or non-epistemic in nature Cohen’s kappa was $\kappa=0.61$, considered to represent *substantial agreement* (El Emam 1999; Fleiss 1971; Landis and Koch 1977; Viera and Garrett 2005). All disagreements were verbally reconciled and the first author coded all remaining justifications. In cases where patterns of disagreement emerged, coding principles were derived and adhered to in scoring the remaining justifications.

Procedure

The order of the study task was as follows. Participants were first provided with directions. Then they accessed a search page that included the first question they were assigned, a textbox to fill in their answer, and the library with hyperlinked buttons from which texts could be

Table 2 Sample response confidence justification coding

Justification coding	Example
Text-directed	Referent to texts in source library
Epistemic	<p>“It was difficult in evaluating whether the source was credible or valid”</p> <p>“I am rather confident because the information...was on [sic. from]multiple sources”</p> <p>“I used a government report for my research and those are considered academic sources.”</p>
Non-epistemic	<p>“Reading the sources provided was hard”</p> <p>“I feel as though I could have written a full paper with the resources provided.”</p>
Task-directed	Referent to question parameters or composed responses
Epistemic	<p>“However, I included in my answer that there was a large variability rate at this age especially between boys and girls, so I think I made it clear it varies between children.”</p> <p>“I only gave estimates with no exact answer.”</p> <p>“I am fairly confident because the information made sense to me”</p>
Non-epistemic	<p>“I’m not really sure it was answering the question”</p> <p>“I was vague in my answer.”</p>
Personally-directed	Referent external to texts or task
Epistemic	<p>“I don’t know much about the topic at hand”</p> <p>“I learned this from class”</p>
Non-epistemic	<p>“I have read and hear a lot about this to begin with.”</p> <p>“I am familiar with this topic and the style of research done in the social sciences. Because of my familiarity and interest, I knew exactly where to look for an answer and how to interpret the information I found.”</p>

accessed. Students could select and use any of the texts in the library while the target question and a textbox was continually displayed. Once students indicated that they had completed their response, by clicking the “next” button, the response confidence page appeared, asking students to rate and justify their response certainty. When students had completed their confidence ratings for the first question answered, indicated by clicking the “next” button, they were presented with the second question, a textbox within which to formulate their response and a library. Once students had completed answering the second question, they were again taken to a page asking them to rate and justify their response confidence.

Results

Overall justifications for response confidence

Across the four questions, students rated their response confidence as 55.95 ($SD=29.84$) on average, on a 100 mm line. In justifying their ratings of response confidence, students produced a total of 940 utterances, with an average of 2.29 justifications offered for each confidence rating ($SD=1.36$). Of the total number of justifications produced, 56.70 % ($n=533$) were text-directed; 18.94 % were task-directed ($n=178$), and 21.38 % ($n=201$) were personally-directed. The remaining justifications (2.98 %, $n=28$), which could not otherwise be classified, were coded as *other*. Of the total number of justifications for response confidence offered, 56.06 % ($n=527$) were epistemic in nature, while 40.96 % ($n=385$) were non-epistemic in nature.

Justifications for response confidence

Our first research question asked: What kinds of justifications do students offer for their response confidence ratings when responding to discrete and open-ended questions in developmental psychology and astrophysics? A repeated measure analysis of variance was used to compare the number of text-directed, task-directed, personally-directed confidence justifications students reported. Mauchly's test indicated that the sphericity assumption had been violated $X^2(2)=26.93$, $p<.001$. Therefore, a Huynh-Feldt adjustment was used ($\epsilon=0.897$). The within-subject effects were overall significant, $F(1,794)=59.34$, $p<0.001$, $\eta^2=0.23$, corresponding to a large effect size. Post-hoc pairwise comparisons, using Bonferroni's adjustment for multiple comparisons with a significance level of $\alpha=0.017$, found that students produced significantly more text-directed justifications for response confidence ($M=2.60$, $SD=2.03$) than both task-directed justifications [$(M=0.87$, $SD=1.40$), $t(204)=9.70$, $p<.001$, Cohen's $d=0.68$] and personally-directed justifications [$(M=0.98$, $SD=1.53$), $t(204)=7.98$, $p<.001$, Cohen's $d=0.56$], in both cases corresponding to a medium effect size. However no significant differences were found between the number of task-directed and personally-directed justification produced, $t(204)=-0.76$, $p=0.45$.

Further, a series of paired sample *t*-tests determined that in providing text-directed justifications for response confidence, students provided significantly more justifications that were epistemic ($M=1.60$, $SD=1.68$) rather than non-epistemic ($M=1.00$, $SD=1.20$) in nature, $t(204)=4.18$, $p<.001$, Cohen's $d=0.29$, corresponding to a small to medium effect size. As a contrast, students provided significantly more task-directed non-epistemic justifications for response confidence, ($M=0.54$, $SD=1.00$), than epistemic justifications ($M=0.33$, $SD=0.81$), $t(204)=2.53$, $p<.05$, Cohen's $d=0.18$. Of the number of personally-directed justifications provided, significantly more were epistemic ($M=0.63$, $SD=1.16$) than non-epistemic ($M=0.35$, $SD=0.97$), $t(204)=2.76$, $p<.01$, Cohen's $d=0.19$. See Table 3 for summary statistics on types of response confidence justifications.

Response confidence across task conditions

Our second research question was: To what extent do students' response confidence ratings and justifications offered differ across question type (i.e., discrete versus open-ended) and academic domain (i.e., developmental psychology versus astrophysics)?

Response confidence ratings There were no significant differences in students' ratings of response confidence across question type (i.e., discrete versus open-ended questions), $t(201)=1.09$, $p=0.28$, as demonstrated via paired sample *t*-test. Students were significantly more confident in their responses to questions in the domain of developmental psychology ($M=60.56$, $SD=27.44$) than in astrophysics ($M=51.66$, $SD=31.48$), $t(201)=3.23$, $p<.01$, Cohen's $d=0.23$.

Justifications for response confidence Justifications for response confidence were examined across question type and domain.

Question type There was a significant difference in the total number of justifications for response confidence students offered when responding to the discrete ($M=2.40$, $SD=1.38$) versus open-ended question ($M=2.19$, $SD=1.33$), $t(204)=2.19$, $p<.05$, Cohen's $d=0.15$.

Table 3 Percent and mean justifications by referent

Justification type	Percent of total	Mean (SD)
Total justifications	100 % (<i>n</i> =940)	4.59 (2.33)
Text-directed	56.70 % (<i>n</i> =533)	2.60 (2.03)
Epistemic	35.00 % (<i>n</i> =329)	1.60 (1.68)
Non-epistemic	21.70 % (<i>n</i> =385)	1.00 (1.20)
Task-directed	18.94 % (<i>n</i> =178)	0.87 (1.40)
Epistemic	7.23 % (<i>n</i> =68)	0.33 (0.81)
Non-epistemic	11.70 % (<i>n</i> =110)	0.54 (1.00)
Personally-directed	21.38 % (<i>n</i> =201)	0.98 (1.54)
Epistemic	13.83 % (<i>n</i> =130)	0.63 (1.16)
Non-epistemic	7.55 % (<i>n</i> =71)	0.35 (0.97)
Discrete	52.34 % (<i>n</i> =492)	2.40 (1.38)
Open-ended	47.66 % (<i>n</i> =448)	2.19 (1.33)
Developmental psychology	51.60 % (<i>n</i> =485)	2.37 (1.25)
Astrophysics	48.40 % (<i>n</i> =455)	2.22 (1.46)

Further, a series of paired sample t-tests revealed significant differences in the types of justifications for response confidence students offered across the two question types. Participants offered significantly more text-directed and text-directed epistemic justifications for response confidence when responding to the discrete question (text-directed: $M=1.51$, $SD=1.30$; text-directed epistemic: $M=0.99$, $SD=1.14$) as compared to the open-ended question (text-directed: $M=1.09$, $SD=1.10$; text-directed epistemic: $M=0.61$, $SD=0.93$); [text-directed: $t(204)=4.61$, $p<.001$, Cohen's $d=0.32$; text-directed epistemic: $t(204)=4.35$, $p<.001$, Cohen's $d=0.30$]. There were no significant differences in the number of text-directed non-epistemic justifications for response confidence reported across question type, $t(204)=0.56$, $p=0.58$.

There were no significant differences in the number of total task-directed justifications, $t(204)=0.14$, $p=0.89$, the number of task-directed epistemic justifications, $t(204)=1.72$, $p=0.09$, nor in the number of non-epistemic task-directed justifications, $t(204)=1.62$, $p=0.11$, students offered when responding to the discrete versus open-ended question.

Students provided both significantly more total personally-directed justifications for response confidence in the case of the open-ended ($M=0.60$, $SD=0.92$) versus the discrete question ($M=0.38$, $SD=0.91$), $t(204)=3.27$, $p<.01$, Cohen's $d=0.23$, and significantly more personally-directed epistemic justifications (discrete: $M=0.24$, $SD=0.63$; open-ended: $M=0.39$, $SD=0.78$), $t(204)=2.63$, $p<.01$, Cohen's $d=0.18$. No significant difference was found in the number of personally-directed non-epistemic reasons students provided, $t(204)=1.58$, $p=0.12$. See Table 4 for response confidence justifications across question types.

Domain Students did not differ in the total number of justifications for response confidence they offered when responding to questions in the two domains, $t(204)=1.49$, $p=0.14$. A series of paired sample t-tests were run to determine if there were domain differences in the types of justifications students offered. First, participants reported significantly more text-directed justifications when substantiating their response confidence ratings for questions in developmental psychology ($M=1.40$, $SD=1.20$) than in astrophysics ($M=1.20$, $SD=1.23$), $t(204)=2.25$, $p<.05$, Cohen's $d=0.16$. Likewise, students offered significantly more text-directed

Table 4 Percent and mean justifications across question type

Justification source Epistemic classification	Question type			
	Discrete		Open-ended	
	M (<i>SD</i>)	Percent	M (<i>SD</i>)	Percent (<i>N</i>)
Total justifications	2.40 ^a (1.38)	100 (<i>n</i> =492)	2.19 ^a (1.33)	100 (<i>n</i> =448)
Text-directed	1.51 ^b (1.30)	62.80 (<i>n</i> =309)	1.09 ^b (1.10)	50.00 (<i>n</i> =224)
Epistemic	0.99 ^b (1.14)	41.26 (<i>n</i> =203)	0.61 ^b (0.93)	28.13 (<i>n</i> =126)
Non-epistemic	0.52 (0.81)	21.54 (<i>n</i> =106)	0.48 (0.74)	21.88 (<i>n</i> =98)
Task-directed	0.043 (0.88)	17.89 (<i>n</i> =88)	0.44 (0.85)	20.09 (<i>n</i> =90)
Epistemic	0.21 (0.65)	8.74 (<i>n</i> =43)	0.12 (0.42)	5.58 (<i>n</i> =25)
Non-epistemic	0.22 (0.58)	9.15 (<i>n</i> =45)	0.32 (0.73)	14.51 (<i>n</i> =65)
Personally-directed	0.38 ^b (0.91)	15.65 (<i>n</i> =77)	0.60 ^b (0.92)	27.68 (<i>n</i> =124)
Epistemic	0.24 ^b (0.63)	10.16 (<i>n</i> =50)	0.39 ^b (0.78)	17.86 (<i>n</i> =80)
Non-epistemic	0.13 (0.61)	5.49 (<i>n</i> =27)	0.21 (0.62)	9.82 (<i>n</i> =44)

^a indicates a significant difference at the $\alpha=0.05$ level

^b indicates a significant difference at the $\alpha=0.01$ level

epistemic justifications after responding to questions in developmental psychology ($M=0.92$, $SD=1.10$; astrophysics: $M=0.68$, $SD=1.00$), $t(204)=2.69$, $p<.01$, Cohen's $d=0.19$). However, no differences across domains were found in the number of text-directed non-epistemic justifications students offered, $t(204)=0.42$, $p=0.68$.

No differences across domain were found in students' task-directed reasons for response confidence; total task-directed reasons, $t(204)=0.56$, $p=0.58$; task-directed epistemic reasons, $t(204)=0.38$, $p=0.70$; or non-epistemic task-directed reasons, $t(204)=0.32$, $p=0.75$.

Finally, in examining students' personally-directed bases for response confidence ratings, no significant differences were found across domains in the total number of such justifications reported, $t(204)=1.57$, $p=0.12$; nor in the number of personally-directed epistemic justifications offered, $t(204)=0.69$, $p=0.49$. Students did differ significantly in the number of personally-directed non-epistemic justifications they cited across the two domains, $t(204)=2.92$, $p<.01$, Cohen's $d=0.20$. Specifically, when responding to questions in the domain of astrophysics students reported more personally-directed non-epistemic reasons for their response confidence ratings ($M=0.25$, $SD=0.69$) than they did when responding to questions in developmental psychology ($M=0.10$, $SD=0.51$). See Table 5 for response confidence justifications across domains.

Response confidence and performance

The third research question in this study examined the correspondence between students' response confidence and types of justifications offered and response accuracy, for the discrete question, and response quality, for the open-ended question. Specifically, we were interested in knowing: In what ways do students' ratings of response confidence and the justifications they provide for those ratings relate (a) to the accuracy of their responses to discrete questions and

Table 5 Percent and mean justifications across domain

	Developmental psychology		Astrophysics	
	Mean (SD)	Percent (N)	Mean (SD)	Percent (N)
Total justifications	2.37 (1.25)	100 ($n=485$)	2.22 (1.46)	100 ($n=455$)
Text-directed	1.40 ^a (1.20)	59.38 ($n=288$)	1.20 ^a (1.23)	53.85 ($n=245$)
Epistemic	0.92 ^b (1.10)	38.97 ($n=189$)	0.68 ^b (1.00)	30.77 ($n=140$)
Non-epistemic	0.48 (0.72)	20.41 ($n=99$)	0.51 (0.84)	23.08 ($n=105$)
Task-directed	0.45 (0.89)	19.18 ($n=93$)	0.41 (0.83)	18.68 ($n=85$)
Epistemic	0.18 (0.56)	7.42 ($n=36$)	0.16 (0.54)	7.03 ($n=32$)
Non-epistemic	0.28 (0.65)	11.75 ($n=57$)	0.26 (0.67)	11.65 ($n=53$)
Personally-directed	0.43 (0.93)	18.35 ($n=89$)	0.55 (0.92)	24.62 ($n=112$)
Epistemic	0.34 (0.79)	14.23 ($n=69$)	0.30 (0.61)	13.41 ($n=61$)
Non-epistemic	0.10 ^b (0.51)	4.12 ($n=20$)	0.25 ^b (0.69)	11.21 ($n=51$)

^a indicates a significant difference at the $\alpha=0.05$ level

^b indicates a significant difference at the $\alpha=0.01$ level

(b) to the characteristics of their responses to open-ended questions in developmental psychology and astrophysics?

Discrete

In responding to the discrete question, 63.41 % of respondents answered correctly ($n=130$). Students did not significantly differ in response accuracy when responding to questions in astrophysics (66.22 %, $n=56$) versus in developmental psychology (64.35 %, $n=74$), $X^2=0.10$, $p=0.75$.

Response confidence ratings Those responding accurately to the discrete questions had significantly higher ratings of response confidence, $F(1, 201)=32.04$, $p<.001$, $\eta^2=0.14$; those answering correctly reported a mean confidence level of 66.01, ($SD=26.52$) as compared to those answering inaccurately ($M=42.23$, $SD=32.81$). A two-way analysis of variance, examining the effects of response accuracy (i.e., correct versus incorrect) and question domain (i.e., astrophysics versus developmental psychology) on response confidence, found only a significant main effect for response accuracy, $F(1, 201)=33.03$, $p<.001$. There was no significant main effect for domain, $F(1, 201)=0.34$, $p=0.56$, and no significant interaction between response accuracy and domain in differentiating response confidence, $F(1, 201)=0.95$, $p=0.33$. The adjusted R^2 for the model was 0.13.

Justifications for response confidence There were no significant differences in the total number of confidence justifications offered by students answering the discrete question accurately versus inaccurately, $F(1, 203)=0.01$, $p=0.92$. Those students responding accurately to the discrete question provided significantly more total text-directed rationales for response confidence (correct: $M=1.76$, $SD=1.29$; incorrect: $M=1.07$, $SD=1.21$), $F(1, 203)=14.47$, $p<.001$, $\eta^2=0.07$. Students responding correctly versus incorrectly did not differ in the number of total task-directed justifications produced, $F(203)=0.93$, $p=0.34$. Students answering the discrete question inaccurately ($M=0.60$, $SD=1.13$) provided significantly more personally-

directed justifications for response confidence than did students answering correctly ($M=0.25$, $SD=0.74$), $F(1, 203)=7.37$, $p<.01$, $\eta^2=0.04$, corresponding to a small effect size.

Open ended

In responding to the open-ended questions, students' answers had an average number of 40.59 words ($SD=43.74$) and students' responses were scored an average of 1.73 ($SD=0.88$) on the SOLO taxonomy, indicating that on average students were attempting to provide multiple pieces of relevant information in their responses. On average, students' responses included 2.94 ($SD=2.05$) pieces of evidence or substantiating claims, and contained an average of 0.80 ($SD=1.18$) elaborations, or explanations of said claims. The mean number of citations students provided in their responses was 0.19 ($SD=0.56$).

A series of independent sample t-tests were run to examine differences in students' performance on the open-ended questions across domains. Students did not significantly differ in their open-ended response word-count, $t(200)=0.06$, $p=0.95$; nor in the number of pieces of evidence, $t(198.28)=1.43$, $p=0.15$; and, elaborations produced when responding to questions in astrophysics vis-à-vis developmental psychology, $t(200)=0.68$, $p=0.50$. However, students did earn higher SOLO scores when responding to open-ended questions in developmental psychology ($M=1.91$, $SD=0.89$) rather than in astrophysics ($M=1.58$, $SD=0.84$), $t(200)=2.71$, $p<.01$, Cohen's $d=0.38$, while providing significantly more citations in the domain of astrophysics ($M=0.27$, $SD=0.68$) than in developmental psychology ($M=0.09$, $SD=0.32$), $t(167.71)=2.53$, $p<.05$, Cohen's $d=0.33$.

Response confidence ratings To examine the effects of open-ended response quality on students' confidence ratings, a multiple linear regression was run predicting open-ended response confidence based on question domain, word count, SOLO score, number of pieces of evidence and elaborations identified, as well as the number of citations included. The model was overall significant, $F(6, 195)=7.94$, $p<.001$, $R^2_{adj}=0.20$. However, only question domain, $\beta=16.53$, $p<.001$, and SOLO score, $\beta=7.46$, $p<.05$ were significant predictors in the model.

Students' reported ratings of response confidence were significantly correlated with a number of measures of response quality, when responding to the open-ended question. Due to data skew and kurtosis, Kendall's Tau was the measure of association used. Specifically, students' ratings of response confidence were significantly correlated with word count [$\tau(202)=0.10$, $p<.05$], SOLO score [$\tau(202)=0.21$, $p<.001$], number of pieces of evidence [$\tau(202)=0.12$, $p<.05$], and number of elaborations [$\tau(202)=0.17$, $p<.01$].

Justification for response confidence The total number of justifications students offered for their response confidence ratings was significantly correlated with all measures of open response quality, including, word count [$\tau(202)=0.33$, $p<.001$], SOLO score [$\tau(202)=0.25$, $p<.001$], the number of pieces of evidence students included [$\tau(202)=0.18$, $p=.001$], the number of elaborations offered [$\tau(202)=0.22$, $p<.001$], and the number of citations included in students' responses [$\tau(202)=0.20$, $p=.001$].

Looking at the number of text-directed reasons for response confidence, the total number of such justifications offered in reference to the open-ended question was significantly correlated with word count [$\tau(202)=0.17$, $p=.001$], SOLO scores [$\tau(202)=0.14$, $p<.05$], the number of reasons students elaborated [$\tau(202)=0.14$, $p<.05$], and the number of citations included in

students' responses [$\tau(202)=0.18, p<.01$]. There were no significant correlations between total text-directed justifications and the number of reasons produced [$\tau(202)=0.10, p=0.09$].

Kendall's Tau correlation analysis determined that the total number of task-directed reasons for response confidence ratings was significantly related to word count [$\tau(202)=0.13, p<.05$], SOLO scores [$\tau(202)=0.12, p<.05$], and the number of reasons included in students' responses [$\tau(202)=0.13, p<.05$], but not significantly correlated with the number of elaborations nor the number of citations in students' answers.

The total number of personally-directed bases for response confidence students reported was not correlated with any measures of open-ended response quality (i.e., word count, SOLO score, pieces of evidence, elaborations, citations).

Discussion

The present study sought to examine students' confidence in their responses to academic tasks requiring the use of multiple texts. This study advances the literature in at least three ways. First, we considered the bases or justifications for students' ratings of response confidence, in addition to their magnitude. Second, we examined the nature of students' response confidence across question types and domains. Finally, we considered students' response confidence ratings and justifications as they related to task performance, all within the context of a multiple text task.

Research question 1: types of justifications

Our first research question examined the types of justifications students offered for reported response confidence ratings. In examining the types of justifications students offered for their ratings of response confidence, we found the majority of these to be text-directed; indicating that students were basing their response confidence ratings on features of texts in the digital library or the information within them. Further, significantly more of these text-directed justifications were epistemic, rather than non-epistemic, in nature. Students reported justifications for response confidence based on general appraisals of text reliability, evaluations of source type and author expertise, and on assessments of the research and evidence cited in texts. This is an encouraging finding for researchers and educators interested in improving students' text evaluation and concerned about the plethora of online texts requiring scrutiny (Braasch et al. 2013; Mason et al. 2010).

In considering the epistemic versus non-epistemic nature of students' justifications for response confidence ratings, we found that while students did attend to epistemic features of texts when justifying their ratings, when providing task-directed justifications, they focused significantly more on non-epistemic features. Indeed, students significantly more frequently evaluated their responses based on superficial features such as their length or their "include[ing] a couple of ideas," rather than considering the quality of evidence in these ideas. Consistent with the MD-TRACE model, this may be interpreted as students evaluating responses in reference to a task model constructed based on limitations in the amount of time or effort students were wanting to expend on the task. When students did justify response confidence in task-directed epistemic ways, they frequently based their confidence on personal convictions or the belief in the accuracy of their responses, without further substantiation. For

example, as students explained, “the factors that I listed I am confident that they are correct,” or, “the answer seems correct.”

At the same time, as suggested by MD-TRACE theory, students ought to have been rendering justifications of response confidence based on task-directed reasoning to a greater extent (e.g., Rouet 2006). While students were not often explicitly comparing the response they produced to a theorized task-model, there were variations in the types of confidence justifications students reported when responding to different types of task.

Research question 2: response confidence across task conditions

Our second research question examined the extent to which students’ response confidence ratings and justifications differed across question type (i.e., discrete versus open-ended) and domain (i.e., developmental psychology versus astrophysics). After responding to the discrete question students generated significantly more text-directed justifications for response confidence than they did after responding to the open-ended question. This may suggest that in evaluating their responses to varied types of questions students were appealing to different task-models.

The task model students constructed was thought to determine the standards they adopted for determining and justifying judgments of response quality. Further, standards for judgments of response quality and how these judgments were justified may have been influenced and constrained by the level of cognitive demand adopting various standards posed. The discrete question was thought to be easier for students to both respond to and to justify their response confidence in; whereas the open-ended question, requiring a more elaborative and integrative response, posed a greater confidence-evaluation challenge for students. This task analysis seems to be consistent with students engaging in the relatively demanding process of response justification based on text-directed factors, but doing so significantly more frequently when responding to the comparatively easier discrete question.

Likewise, text-directed justifications for response confidence were reported significantly more often when students responded to questions in developmental psychology, the domain they were familiar and experienced in, rather than to questions in astrophysics. We thought this pattern emerged because developmental psychology questions constituted a more familiar task for our participants. That is, when answering questions in the comparatively more familiar domain (i.e., developmental psychology), students could engage in the relatively demanding text evaluation process and offer significantly more text-directed justifications for response confidence without making their task excessively difficult.

The finding that students offered significantly more text-directed justifications when rating their confidence in responses to developmental psychology questions was contrary to our expectations. Since our participants were primarily social science majors, we expected these students to have more background in developmental psychology than in astrophysics, and therefore to be more text-directed in justifying their response confidence ratings in the latter. We thought that, when responding to questions in an unfamiliar domain (i.e., astrophysics), participants would need to rely on text-based information to substantiate their answers to a greater extent. Whereas, in responding to questions in the domain of developmental psychology, we expected students to be more personally-directed, having greater stores of prior knowledge and experience to draw on

when justifying their ratings of response confidence. The fact that our data indicate the opposite justification pattern may be a testament to the relative difficulty students have with text evaluation as a contributory factor in their formulation of text-directed justifications for response confidence. Students were able to engage in the text based justification process significantly more often when they presumably had familiarity with and prior knowledge in a domain. Indeed, the importance of learner factors, such as prior knowledge, in students' text processing is a finding well-established in prior research (e.g., Alexander et al. 1994; Bråten et al. 2011).

Students reported significantly more personally-directed justifications for response confidence when answering open-ended questions. This finding may be associated with the higher level of difficulty these more elaborative questions pose for students. One strategy participants might have employed to cope with the greater complexity introduced by open-ended questions was to reformulate such tasks from ones requiring the reconciliation and integration of evidence to ones requiring simply a relevant opinion, where any answer may be as good as any other. This is a response we have identified in our prior research (List et al. 2013). In effect, this strategy might specifically serve to mitigate the challenge posed by open-ended questions having flexible and varied criteria for what may constitute a quality response. As suggested by the MD-TRACE model, this may result in students' difficulties building a task model and evaluating open-ended responses in reference to said model. Participants considering open-ended questions to be questions of opinion would thus, indeed, turn to personally-directed justifications for response confidence, convinced that they know a potential answer, without considering whether their personal knowledge would lead to an answer of quality or the best answer.

When responding to both the discrete and open-ended questions, students frequently justified response confidence based on their finding "the answer" stated directly within a text. These included explanations for response confidence ratings such as, "The answers to the questions were easy to find," and, "Many of the sources had varying answers and none of them explicitly stated the answer to the question." In our study, these were coded as text-directed as students were justifying response confidence based on information found in texts or lack thereof. However, justifications such as these also provide insights into students' interpretations of task demands. A number of our participants seemed to want responses that were stated explicitly in texts and to be able to find these answers with speed and ease. While in the case of the discrete question, we would expect students to locate a specific amount reported in the texts and potentially verify it across other texts, in the case of the open-ended question, we expected participants to gather, evaluate, reconcile, and integrate information provided in various texts in composing an elaborative response. Such a response process has been outlined in theoretical models of multiple source use (e.g., Britt et al. 1999; Rouet 2006). Yet, repeatedly the students in our study either believed that, as with the discrete question, an adequate open-ended response would be explicitly stated in a text and tried to identify this text or they were frustrated that they were unable to find a directly stated answer to the open-ended questions. One student explained her low rating of confidence in her response to the open-ended question by saying the "answer wasn't displayed out in the open." Justifications such as this require us to examine more closely how students perceive the assignments they are given and which tasks students carry out when composing responses based on multiple texts.

Research question 3: response confidence and performance

In response to our third research, justifications for response confidence, in the case of both the discrete and open-ended questions, were tied to performance. As expected, those participants responding accurately to the discrete question were significantly more likely to cite text-directed justifications, indicating that these students were locating their answers in text and potentially corroborating their answers across multiple texts. Interestingly, those participants responding incorrectly more often justified their response confidence in ways that were personally-directed. Such justifications were reliant on students believing they knew the correct answer to the question from prior knowledge or experience. Convictions that proved incorrect. This finding is consistent with prior research indicating that generally students are often poor at judging what they know (Dinsmore and Parkinson 2013).

For the open-ended question, indicators of response quality were significantly correlated with the total number of text-directed justifications and task-directed justifications for response confidence reported. Indicators of response quality were not significantly correlated with students' personally-directed justifications, suggesting that while participants may have thought that they had sufficient knowledge or experience to generate an acceptable answer to the open-ended questions, they may not have. Although statistically significant, these correlations showed only a limited association between open-ended measures of response quality and justifications for response confidence. Further work is needed to better understand the nature of these associations.

There is a need to examine students' response confidence further. Evidence from this study suggests that it is a fruitful pursuit. Participants in our study demonstrated not only an ability to make judgments of response confidence, but also to justify these ratings along varying dimensions. However, more work is needed to understand the text, task, and personal factors that may impact response confidence and more importantly, the relation between response confidence ratings and subsequent multiple source use behaviors when completing academic tasks.

Limitations

A number of limitations in study design should be considered, including those associated with sample selection, task formulation, justification coding, and text presentation. First, the domain-based differences identified in students' response confidence ratings and justifications are conflated with students' varied experiences in these domains. Our participants had greater familiarity and prior knowledge of developmental psychology. Although, as expected, students expressed greater confidence in their developmental psychology responses, this was likely due to their experience in the domain, rather than characteristics of the domain. Further research with cross-sectional populations is needed to explore domain differences in students' response confidence and confidence justifications.

Further, more work is needed to clarify which aspects of assigned questions (i.e., discrete and open-ended) were associated with differences in confidence judgments and justifications. The discrete and open-ended questions presented varied with regard to both the nature of the response demanded and the nature of the scoring. Discrete

questions required both a singular answer and were scored as correct or incorrect, while open-ended questions allowed for a more elaborated response that was also coded in a less rigid fashion, with no correct answer. Future research should be directed at disentangling these factors. For example, insights may be gained from examining how response confidence and justifications vary when students, regardless of the level of elaboration required, are assigned questions as having a single correct answer or rigid response standard.

Coding the justifications participants provided for response confidence was a subjective process. While the coding scheme used was able to reliably classify the majority of students' justifications, the coding conventions adopted were reflected in the results. For instance, defaulting to coding justifications that considered aspects of both text and task as text-directed, contributed to the finding that the majority of justifications students provided were text-directed. More work is needed to determine the reliability of the justification coding scheme used across additional MSU tasks.

In the present study, texts were presented to students as an array of horizontal buttons in alphabetical order, by source type, with no additional document information included. This delimitation was adopted for two primary reasons. When presented with texts in a list-like fashion, resembling a search engine results page, students have been found to have a strong order bias in selecting those texts appearing at the top (Kammerer and Gerjets 2010) and to select texts based on superficial relevance cues (Rouet et al. 2011), rather than quality indicators. In this study, by presenting texts according to source type, we hoped to draw students' attention to a key document features associated with judgments of text trustworthiness (Bråten et al. 2009; Rouet et al. 1996). However, this form of presentation came at the cost of ecological validity and may have, in part, increased students' reporting of text-directed justifications. Source type was considered to be a salient document feature upon which students could base their text selections. Indeed, studies have shown students to evaluate texts based on source type based schema (Bråten et al. 2011). Yet, providing additional document information (e.g., author credentials) may have aided students in their text selections and evaluations and in forming text-directed justifications for response confidence. Future research should consider whether presenting additional information about texts introduces more text-directed criteria into students' justifications for response confidence.

Future directions

In this study, students were asked to rate their response confidence following task completion, after answers to each question had been submitted. However, there is a need to understand how students might modify their text use behaviors or responses if asked to evaluate their response confidence at various stages throughout the multiple source use process. Further, in this study, students were prompted to both rate and justify their response confidence. Self-evaluating answers has been identified as a high-level regulatory strategy and as such it is important to investigate not only whether students are able to engage in response evaluation, but also the extent to which they do so spontaneously when completing tasks. Finally, the task instructions provided to students in this study asked respondents to answer each questions as they would for an academic class, leaving the criteria along which each response would be evaluated intentionally vague. This was done to understand the task-directed criteria students would self

Appendix

Source type	Discrete astrophysics			Open-ended astrophysics			Discrete developmental psychology			Open-ended developmental psychology		
	Title and source	Word Count	Flesch Kincaid readability and grade level	Title and source	Word count	Flesch Kincaid readability and grade level	Title and source	Word count	Flesch Kincaid readability and grade level	Title and Source	Word Count	Flesch Kincaid Readability and Grade Level
Government Report	How Many Stars Are There in the Universe? <i>European Space Agency</i>	552	59.8/9.8	Was Venus once a habitable planet? <i>European Space Agency</i>	554	50.4/10.5	Validation of National Child Language Measures at 14 and 24 months <i>Administration for Children and Families</i>	991	55.8/8.5	Early Child Development <i>World Health Organization</i>	829	34/13.6
Journal Article	RAVE: The Radial Velocity Experiment <i>GaIA Spectroscopy, Science and Technology</i>	1281	45.1/11.7	A Mission to Titan, Europa, or Enceladus? <i>Cosmology</i>	3708	37.1/13	Productive Vocabulary Size Development in Children Ages 18-24 months: Gender Differences <i>Speech, Music, and Hearing</i>	1854	54.7/9.5	The Role of Home Literacy Practices in Preschool Children's Language and Emergent Literacy Skills <i>Journal of Speech, Language, and Hearing Research</i>	1736	27.2/16.9
Newspaper	Cast Adrift in the Milky Way, Billions of Planets, All Alone <i>New York Times</i>	822	51.3/12	New Planet May Be Able to Nurture Organisms <i>New York Times</i>	971	63.4/9.3	Child Development: Language Takes on New Significance <i>New York Times</i>	400	46.5/12.2	RX: Read to Your Baby <i>New York Times</i>	1399	63.7/8.2
News Site	Hyperfast Star Was Booted from Milky Way <i>Science Daily</i>	1084	58.7/10.2	Newly Discovered Planet May be First Truly Habitable Exoplanet <i>Science Daily</i>	1077	48.8/11.7	Young Toddlers Think In Terms of Whole Object, Not Just Parts <i>Science Daily</i>	415	57.6/10.5	Psychologists Test Effects of Household Noise on Children's Verbal Development <i>Science Daily</i>	689	67.1/8.9
Online Encyclopedia	Milky Way <i>Wikipedia</i>	694	59.3/10.3	Planetary Habitability <i>Wikipedia</i>	1390	26.2/14.7	Vocabulary Development	1180	56.1/9.1	Language Development <i>Wikipedia</i>	814	38.3/12.8

(continued)

Source type	Discrete astrophysics			Open-ended astrophysics			Discrete developmental psychology			Open-ended developmental psychology		
	Title and source	Word Count	Flesch Kincaid readability and grade level	Title and source	Word count	Flesch Kincaid readability and grade level	Title and source	Word count	Flesch Kincaid readability and grade level	Title and Source	Word Count	Flesch Kincaid Readability and Grade Level
Blog	Image of the Day: A Billion Stars of the Milky Way <i>The Daily Galaxy</i>	505	34.8/16.4	Presence of Life May Be Required to Maintain Habitability of a Planet over Billions of Years <i>The Daily Galaxy</i>	220	22.9/16.2	<i>Wikipedia</i> Two Years <i>Wonder Time</i>	545	61.8/9.4	Word by Word <i>Wonder Time</i>	2540	65/9.6
Website	How Many Stars are in the Milky Way? <i>Universe Today</i>	408	71.2/8.8	Habitable Planet <i>Universe Today</i>	420	45.2/11.6	Toddler Milestones: Talking <i>Baby Center</i>	907	72.4/7.6	Your 24-month-old's Learning and Cognitive Development: Asking Questions <i>Baby Center</i>	595	62.3/9.5

generate in evaluating their responses, however, future work should consider students' response confidence ratings when participants are provided with more specific criteria for response formulation.

References

- Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts. *Review of Educational Research*, *64*(2), 201–252. doi:10.3102/0034654306002201.
- Alexander, P. A. (2006). *Psychology in learning and instruction*. Upper Saddle River, NJ: Pearson.
- Biggs, J. B. (1979). Individual differences in study processes and the quality of learning outcomes. *Higher Education*, *8*(4), 381–394.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic.
- Braasch, J. L. G., Bråten, I., Strømso, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, *38*(3), 180–195. doi:10.1016/j.cedpsych.2013.03.003.
- Bråten, I. (2008). Personal epistemology, understanding of multiple texts, and learning within Internet technologies. In M. S. Khine (Ed.), *Knowing, knowledge, and beliefs: Epistemological studies across diverse cultures* (pp. 351–376). Amsterdam: Springer.
- Bråten, I., Strømso, H. I., & Samuelstuen, M. S. (2008). Are sophisticated students always better? The role of topic-specific personal epistemology in the understanding of multiple expository texts. *Contemporary Educational Psychology*, *33*(4), 814–840.
- Bråten, I., Strømso, H. I., & Britt, M. A. (2009). Trust matters: examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*, 6–28. doi:10.1598/RRQ.44.1.1.
- Bråten, I., Strømso, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, *21*(2), 180–192.
- Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. Van Den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–234). Mahwah: Lawrence Erlbaum Associates.
- Currie, L., Devlin, F., Emde, J., & Graves, K. (2010). Undergraduate search strategies and evaluation criteria: searching for credible sources. *New Library World*, *111*(3/4), 113–124.
- Dinsmore, D. L., & Parkinson, M. P. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, *24*, 4–14.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280.
- El Emam, K. (1999). Benchmarking Kappa: inter-rater agreement in software process assessments. *Empirical Software Engineering*, *4*(2), 113–133.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, *59*(4), 395–430.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382.
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømso, H. I. (2010). Summary versus argument tasks when working with multiple documents: which is better for whom? *Contemporary Educational Psychology*, *35*(3), 157–173.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory and Cognition*, *15*(1), 84–93. doi:10.3758/BF03197714.
- Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Faris & D. Bloom (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 317–351). Greenwich: Information Age Publishing.
- Goldman, S. R., & Scardamalia, M. (2013). Managing, understanding, applying, and creating knowledge in the information age: next-generation challenges and opportunities. *Cognition and Instruction*, *31*(2), 255–269. doi:10.1080/07370008.2013.773224.
- Goldman, S. R., Braasch, J. L., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from Internet sources: processing patterns of better and poorer learners. *Reading Research Quarterly*, *47*(4), 356–381.

- Greene, J. A., Muis, K. R., & Pieschl, S. (2010). The role of epistemic beliefs in students' self-regulated learning with computer-based learning environments: conceptual and methodological issues. *Educational Psychologist, 45*(4), 245–257.
- Grimes, P. W. (2002). The overconfident principles of economics student: an examination of a metacognitive skill. *The Journal of Economic Education, 33*(1), 15–30.
- Grimes, D. J., & Boening, C. H. (2001). Worries with the web: a look at student use of web resources. *College and Research Libraries, 62*(1), 11–23.
- Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: examining the nature of judgments of confidence. *Learning and Instruction, 24*, 37–47.
- Hartley, K., & Bendixen, L. D. (2001). Educational research in the internet age: examining the role of individual characteristics. *Educational Researcher, 30*(9), 22–26.
- Kammerer, Y., & Gerjets, P. (2010). How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface. Proceedings from ETRA 2010: *Symposium on Eye-Tracking Research and Applications* (pp. 299–306). Association for Computing Machinery
- Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119*(1), 80–113. doi:10.1037/a0025648.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: the place of response certitude. *Educational Psychology Review, 1*(4), 279–308. doi:10.1007/BF01320096.
- Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1979). Feedback and content review in programmed instruction. *Contemporary Educational Psychology, 4*(2), 91–98.
- Kulhavy, R. W., Stock, W. A., Thornton, N. E., Winston, K. S., & Behrens, J. T. (1990). Response feedback, certitude, and learning from text. *British Journal of Educational Psychology, 60*(2), 161–170. doi:10.1111/j.2044-8279.1990.tb00933.x.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
- Lenski, S. D. (1998). Intertextual intentions: making connections across texts. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas, 72*(2), 74–80. doi:10.1080/00098659809599601.
- List, A., Grossnickle, E. M., & Alexander, P. A. (2013). Examining epistemic and nonepistemic dimensions of source evaluation. In M. T. McCrudden (Chair), *Google we trust: Examining evaluations of source and search results*. Symposium conducted at the annual meeting of the American Educational Research Association, San Francisco.
- Maki, R. H., & Serra, M. (1992). The basis of text predictions for text material. *Journal of Experimental Psychology, 18*(1), 116–126.
- Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*(4), 723–731.
- Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: are students epistemically active? *Instructional Science, 38*(6), 607–633. doi:10.1007/s11251-008-9089-y.
- Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction, 21*(1), 137–151.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*(1), 57–86.
- Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*(3), 271–290.
- Moore, D. R. (2007). Concept acquisition and confidence using a spatial probability measure instrument. *Journal of Educational Multimedia and Hypermedia, 16*(1), 25–38.
- Mosenthal, P. B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Educational Research Journal, 35*(2), 269–307.
- Muis, K. R. (2007). The role of epistemic beliefs in self-regulated learning. *Educational Psychologist, 42*(3), 173–190.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the "delayed-JOL effect". *Psychological Science, 2*(1991), 267–270. doi:10.1111/j.1467-9280.1991.tb00147.x.
- Pieschl, S. (2009). Metacognitive calibration: an extended conceptualization and potential applications. *Metacognition and Learning, 4*(1), 3–31.
- Prensky, M. (2001). Digital natives, digital immigrants: part 1. *On the Horizon, 9*(5), 1–6.
- Pressley, M., Ghatala, E. S., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly, 25*(3), 232–249.

- Purcell, K., Rainie, L., Heaps, A., Buchanan, J., Friedrich, L., Jacklin, A., Zickuhr, K. (2012). *How teens do research in the digital world*. Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/~media/Files/Reports/2012/PIP_TeacherSurveyReportWithMethodology110112.pdf. Accessed 1 Nov 2013.
- Rouet, J.-F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah: Lawrence Erlbaum Associates.
- Rouet, J.-F., & Britt, M. A. (2011). Relevance processing in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Greenwich: Information Age Publishing.
- Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, *88*(3), 478–493.
- Rouet, J.-F., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction*, *21*(2), 205–219.
- Salmerón, L., & Kammerer, Y. (2012). Selecting pages from Google to learn about a controversial topic: The role of epistemic belief. In E. de Vries & K. Scheiter (Eds.), *Proceedings EARLI special interest group text and graphics: Staging knowledge and experience: How to take advantage of representational technologies in education and training?* (pp. 181–183). Grenoble: Université Pierre-Mendès-France.
- Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). New York: Routledge.
- Stahl, S., Hynd, C., Britton, B., McNish, M., & Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly*, *31*(4), 430–456. doi:10.1598/RRQ.31.4.5.
- Strømso, H. I., Bråten, I., & Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and Instruction*, *18*(6), 513–527. doi:10.1016/j.learninstruc.2007.11.001.
- Vasilyeva, E., De Bra, P., Pechenizkiy, M., Puuronen, S. (2008). *Tailoring feedback in online assessment: influence of learning styles on the feedback preferences and elaborated feedback effectiveness*. Paper presented at Eighth IEEE International Conference, 1–5, 834–838.
- Vasilyeva, E., Pechenizkiy, M., & De Bra, P. (2008b). Adaption of elaborated feedback in e-learning. In W. Nejdl et al. (Eds.), *Proceedings of the international conference on adaptive hypermedia (AH'08)* (Lecture Notes in Computer Sciences, Vol. 5149, pp. 235–244). Berlin: Springer.
- Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology*, *102*(4), 817.
- Viera, A. J., & Garrett, J. M. (2005). Understanding inter-observer agreement: the kappa statistic. *Family Medicine*, *37*(5), 360–363.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, *52*(1), 234–246. doi:10.1016/j.compedu.2008.08.003.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah: Lawrence Erlbaum Associates.