APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **18**: 203–221 (2004)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.959

Prior Beliefs and Methodological Concepts in Scientific Reasoning

ANDREA FOLLMER GREENHOOT*, GEORGE SEMB,
JOHN COLOMBO and THOMAS SCHREIBER

University of Kansas, USA

SUMMARY

This investigation focuses on the relations between prior beliefs, methodological concepts, and college students' ($N = 211$) scientific reasoning in different problem contexts. Participants were presented with exercises that described the method and results of experiments, and were asked to draw conclusions about the causal status of variables that violated their prior beliefs. Participants drew conclusions in both abstract (i.e. recommend a conclusion for the fictional experimenter), and personal settings (i.e. draw their own conclusions about the phenomenon). Participants' recommendations for the hypothetical experimenter were predicted by their understanding of two methodological concepts: the function of empirical evidence and experimental control. Students' personal conclusions were predicted by their prior beliefs and their appreciation of the objectivity of inquiry. Thus, even when students understand in the abstract how data should relate to scientific conclusions, prior beliefs often take precedence in their own conclusions, especially when they do not understand the issue of objectivity. Copyright © 2004 John Wiley & Sons, Ltd.

Recent interdisciplinary trends in the field of cognitive psychology have emphasized the importance and relevance of the study of cognition for understanding how people learn in educational and instructional contexts (Bransford, Brown, & Cocking, 1999). One of the areas in which the interface between cognition and education is particularly salient is the development of critical thinking skills inherent to scientific reasoning. In this realm, information is presented to students with the goal of having them use empirical algorithms to analyse and solve a wide array of problems. Successful execution of such scientific reasoning requires that students regard the empirical method as a reliable source of knowledge, irrespective of their own biases toward phenomena and events in their world. Educators view the development of these skills as highly desirable in all students, even those with no interest in scientific careers, because they should make students effective decision makers and problem solvers in their own lives (e.g. Bransford et al., 1999; Byrnes, 2001). Yet considerable research suggests that students may not always make use of their abilities in academic or real-life settings (Klaczynski, 2000, 2001; Perkins, Jay, & Tishman, 1993; Stanovich & West, 1998). To address these issues in the current study, we examine college students' scientific reasoning about data that are likely to conflict with

*Correspondence to: Dr Andrea Follmer Greenhoot, Department of Psychology, 1415 Jayhawk Blvd., 426 Fraser Hall, University of Kansas, Lawrence, KS 66045, USA. E-mail: agreenhoot@ku.edu

their biases or prior beliefs, as well as the factors that predict high quality versus sub-optimal reasoning. We were interested not only in students' abilities to draw accurate conclusions based on the data, but also in the likelihood of personal belief change in response to contradictory evidence. This work was motivated both by theoretical questions about the factors involved in high quality scientific reasoning and by practical questions about educational strategies for teaching scientific content and scientific reasoning processes more effectively.

Much of the recent literature on scientific reasoning focuses on the acquisition of core methodological concepts thought to be fundamental to scientists' causal reasoning. These concepts include, among others, an appreciation of (a) the function of empirical evidence (Kuhn, Amsel, & O'Laughlin, 1988), (b) objectivity of inquiry (Klahr, Fay, & Dunbar, 1993; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Penner & Klahr, 1996; Schauble, 1990, 1996), (c) reliability (Schauble, 1996), and (d) experimental control and covariation (Chen & Klahr, 1999; Koslowki, Okagaki, Lorenz, & Umbach, 1989; Kuhn et al., 1995; Schauble, 1996). Among college students there appears to be considerable variation in the appreciation of these concepts. For example, Koslowski et al. (1989) reported that both children and adults tended to disregard information about experimental control when asked to make cause and effect conclusions about data. Similarly, Schauble (1996) found that neither children nor adults considered the possibility of measurement error (i.e. reliability) when asked to draw conclusions from their own physical measurements. Although these methodological concepts seem to be widely accepted as fundamental to the scientific reasoning process, the current literature provides little empirical evidence about how appreciation of these concepts actually relates to the quality of scientific reasoning in different settings. To address this issue in the current study, we look at whether comprehension of these basic concepts increases the likelihood that students draw accurate conclusions about different kinds of scientific phenomena.

It is also clear that high quality scientific reasoning involves far more than the simple application of these methodological concepts and inductive reasoning to examine data and make valid conclusions (Koslowki et al., 1989; Kuhn et al., 1995; Schauble, 1996). Rather, the critical reasoning skills that characterize the empirical method often involve the suspension, or perhaps even the rejection, of one's prior beliefs about phenomena or events in the face of contradictory evidence. Some have argued that the ability to draw accurate conclusions when data conflict with prior beliefs requires a form of metacognitive reasoning, sometimes referred to as 'metacognitive distancing' (e.g. Kuhn et al., 1995; Sigel, 1993). Similarly, Stanovich (e.g. 1999) has described the ability to reason independently of prior knowledge or beliefs as 'decontextualized thinking.' A considerable literature, however, suggests that metacognitive distancing or decontextualized thinking is quite difficult; prior beliefs have been shown to bias all sorts of cognitive operations, including memory, impression formation, reasoning, and problem solving (e.g. Klaczynski, Gordon, & Fauth, 1997; Ross, 1989; Stangor & McMillan, 1992; Stanovich & West, 1998). Research on scientific reasoning illustrates that both children and adults can be influenced by their own prior beliefs about a phenomenon, particularly when their domain-relevant knowledge is low. When confronted with data that contradict prior beliefs, people are often reluctant to relinquish them or to make major changes to their underlying theory about the phenomenon (e.g. Klahr & Dunbar, 1988; Klayman & Ha, 1987; Kuhn et al., 1988; Schauble, 1996). Similarly, research on various types of scientific misconceptions illustrates that such misconceptions are highly resistant to change through traditional instruction (e.g. Chinn & Brewer, 1993; Winer, Cotrell, Gregg, Fournier, & Bica, 2002). Even when students provide the correct answers about commonly misunderstood phenomena immediately after instruction, their misconceptions often reappear when they are questioned weeks or months later (Bransford et al., 1999; Winer et al., 2002). One interpretation of these patterns is that the students' own personal beliefs about the phenomenon were not altered by their educational experiences.

Although it is clear that prior beliefs can bias one's scientific conclusions, less is known about the conditions under which students will suspend or relinquish prior beliefs in response to data that violate expectations. Research on critical thinking and decision-making suggests that reasoning biases are predicted by both task characteristics and individual differences variables such as general cognitive ability and a flexible, open-minded thinking style (Klaczynski et al., 1997; Sa, West, & Stanovich, 1999; Stanovich & West, 1997, 1998). In the current study, individual differences in students' appreciation of methodological concepts might predict the likelihood that they will come to accurate scientific conclusions despite contradictory beliefs. For example, students with greater appreciation of the function of empirical evidence should be more likely to attend to data, and therefore more likely to draw data-based conclusions, than students with a lesser understanding of this concept. Indeed, a recent study by Leshowitz and his colleagues indicated that students who were instructed on methodological criteria or rules for evaluating evidence showed less biased reasoning about the results of scientific studies than students who did not receive such instruction (Leshowitz, DiCerbo, & Okun, 2002). Students might also reason differently about scientific phenomena in different contexts. Specifically, students might have greater difficulty overcoming their biases when asked to draw their own personal conclusions than when asked to reason in a more abstract situation. In other words, the roles of methodological concepts and prior beliefs in scientific reasoning might vary depending on whether students are reasoning in an abstract context or from a more personal perspective.

In the current investigation we examine the links between methodological concepts, prior beliefs, and scientific reasoning in different contexts. The specific aims were to: (1) examine the role of core methodological concepts in college students' reasoning about data that are likely to contradict their prior beliefs, (2) determine whether students have greater difficulty overcoming contradictory prior beliefs when asked to come to their own personal conclusions than when asked to reason in a more abstract situation, and (3) determine whether students' reliance on core methodological concepts differs between formal, abstract settings and real-world or personal contexts that require scientific thinking. We conducted a study of more than 200 university undergraduates' scientific reasoning under conditions in which their a priori biases toward particular scientific phenomena were likely to be correct or incorrect. We first assessed their appreciation of four methodological concepts in the context of a series of hypothetical experiments. We then examined how participants' understanding of these concepts was related to the quality of their performance on exercises designed to assess scientific reasoning. Each exercise described the method and results of an experiment evaluating the effects of two target variables on a particular outcome, one in which participants' prior beliefs were likely to be correct, and one in which their prior beliefs were likely to be incorrect. Students were asked to identify the appropriate conclusions for the experimenter who had conducted the investigation (the abstract setting) and to draw their own conclusions about the effects of the factors on the outcome (the personal setting).

METHOD

Participants

A total of 419 students enrolled in an introductory child development course at the University of Kansas during the Spring 2001 semester were invited to participate in the study. This survey course is typically used to fulfil a social science requirement in the College of Liberal Arts and Sciences. There were 188 freshmen, 113 sophomores, 54 juniors, and 64 seniors. Ninety-nine students were male and 320 were female. For 209 students it was a required course, and for 210 it was an elective that fulfilled general liberal arts requirements. Thus, this sample provides a fair cross-section of the undergraduate population at this institution. Out of 419 students, 407 consented to participate in the study. Participating students reported a mean grade point average (GPA) of 2.96 on a 4-point scale.

Of the 407 students who agreed to participate, 235 completed both scientific exercises and the reading test that comprised the battery of variables for the study. Only students who had completed all three of these measures were included in the analysis. An additional 24 students were excluded because they failed to report GPA, which was a control variable included in our analyses. Thus, the present analysis was based on 211 participants.

Materials

Science exercises

There were two science exercises. Both problems followed a format in which students were asked to reason about the effects of two target variables on a particular dependent variable. One problem (the Ball problem) involved predicting how the speed of a ball rolling down a ramp would be affected by (1) the angle of the ramp's incline, which does affect the dependent variable, and (2) the weight (mass) of the ball, which does not affect the dependent variable. The second problem (the Language problem) asked participants to reason about the effect on language development on (1) how much parents read to their children, which has been consistently linked to improved language outcomes (see e.g. Wells, 1985a, 1985b) and (2) the amount of television children watch, which is popularly believed to have negative effects on language outcomes (Shannon & Fernie, 1985), but on which point the scientific data suggest equivocal effects (e.g. Rice, 1983; Singer & Singer, 1983; Wright et al., 2001). Both exercises may be viewed at www.people.ku.edu/~dea.

Each problem had three sections. The first segment of each exercise comprised the *Prior Beliefs* section, in which students indicated their existing knowledge or beliefs about the effect of each factor on the outcome. This section was followed by a section that assessed their understanding of *Methodological Concepts*, which presented descriptions of several investigative methods employed by fictional 'experimenters' to determine the effects of each target variable on the outcomes. These hypothetical scenarios varied in terms of their use of evidence-based methods, the reliability of the measurements, experimental strategies, and the influence of prior beliefs. Participants' responses to questions about the efficacy of these different investigative methods yielded measures of their understanding of four methodological concepts. Finally, to evaluate scientific reasoning, in the *Conclusions* section participants were presented the method and results of a scientifically sound, hypothetical experiment and asked to draw conclusions about the effects of each of the two variables on the outcome.

For each exercise, the 'observed' or reported effect of one of the target factors violated most adults' prior knowledge or beliefs about that factor. Because the results were unlikely to correspond with participants' existing beliefs about the phenomenon in question, we labelled this variable the 'Belief-Inconsistent' factor. For the Ball exercise, the described manipulations of ball weight had no effect on ball speed, even though most adults believe that ball weight would be positively related to speed. Thus, the Belief-Inconsistent factor for this domain was ball weight. For the Language problem, the Belief-Inconsistent factor was television viewing. Contrary to most adults' beliefs, the manipulation of television viewing frequency did not affect language development in our hypothetical experiment. The effect of the other factor (the 'Belief-Consistent' factor) in each exercise was intuitive or consistent with common knowledge. The Belief-Consistent factor in the Ball problem was angle of incline; the observed results of the experiment indicated that angle size was positively related to ball speed. Similarly, for the Language problem, the Belief-Consistent factor, reading frequency, was positively related to language development.

Prior beliefs. The first two questions of each exercise measured participants' prior beliefs about the scientific problems. For each problem, students were asked to indicate how they thought each factor was related to the outcome. To illustrate, on the Ball problem, students were asked, 'How do you think the speed of the ball down a ramp is affected by the angle of incline?' Possible responses included a positive effect ('The steeper the angle, the faster the ball will roll down the ramp'), a negative effect ('The steeper the angle, the slower the ball will roll down the ramp') or no effect ('The angle of the ramp is not related to how fast the ball will roll down the ramp'). Students' responses to each question were scored as accurate or inaccurate, depending on whether they correctly predicted the relationship between the variable and the outcome (according to the results of the hypothetical experiment later described to them). For the Ball problem, the response, 'the steeper the angle of incline, the faster the ball will roll' was scored as accurate, and all other responses were scored as inaccurate. For the weight dimension of this problem, the accurate response was, 'the weight of the ball is unrelated to how fast the ball will roll.' Likewise, for the Language problem, predictions of a positive association between reading and language development were coded as accurate, as were predictions of no association between television-watching and language development.

Methodological concepts. The Methodological Concepts section measured students' understanding of four scientific concepts: function of evidence, reliability, objectivity, and experimental control. Participants were presented a series of scenarios that featured experimenters attempting to address the primary questions in each domain with different methods. The respondents were asked to answer questions about the effectiveness of these different methods. Measures of each concept were constructed by pooling responses from several items that addressed the concept. Many of the raw items used a 7-point rating scale that asked students to 'strongly agree' or 'strongly disagree' with a statement about the concept, to rate the likelihood of an outcome on a 7-point scale from 'highly unlikely' to 'highly likely,' or to indicate their confidence in a statement on a 7-point scale from 'not at all confident' to 'highly confident.' A 7-point scale was selected so as to give participants the opportunity to take a completely neutral position on these items. Some items required a discrete multiple-choice response that was scored as either correct or incorrect.

Function of evidence. An appreciation of the function of evidence involves a view of empirical data as the most fundamental standard for evaluating any claim, hypothesis, or belief about a phenomenon. To measure participants' understanding of the function of evidence, they were asked to indicate their confidence in scientific conclusions based on empirical investigations (i.e. conclusions of a student who has conducted an experiment) and non-empirical methods (e.g. conclusions of a student who had prior indirect experience with similar situations, or a student who merely 'thought hard' about the problem). Two questions asked about empirical methods. One of these questions asked participants to indicate their confidence in the conclusions of a student who conducted an experiment, whereas the other question asked them to indicate their level of agreement with a general statement about the importance of gathering evidence to solve scientific problems. A mean confidence rating for evidence-based methods was calculated as the average of the students' responses to these two questions. Four questions focused on non-empirical methods. Participants indicated their confidence in the conclusions of a student who solved the problem by thinking about prior indirect experience with similar situations, and in the conclusions of a student who merely 'thought hard' about the problem. They also indicated their agreement with general statements about the ability to solve scientific problems using logical thinking or prior experience. Participants' responses to these four questions were averaged together to yield a mean rating for non-empirical methods. A principal components analysis on all six questions confirmed that these items formed two separate factors. For each participant, an overall Function of Evidence score was calculated as the difference between his or her mean ratings of empirical methods and of non-empirical methods. Higher scores on this variable indicated a greater appreciation of the function of empirical evidence.

Reliability. Judgments about reliability necessitate an understanding and evaluation of the quality of measurement exercised in collecting the evidence. The ability to make such judgments leads to consideration of the number, integrity, and method of measurements made in the collection of empirical evidence. Two questions in our exercises measured participants' understanding of reliability. The first question focused on measurement reliability and asked students to rate the likelihood, on a scale of 1 to 7 (with 7 being 'highly likely') that a hypothetical investigator would get the same measurement of ball speed or language development (number of words used) twice. The second question focused on interrater reliability. Using the same scale, participants rated the likelihood that two hypothetical experimenters would get identical measurements of ball speed or word use. For each student, an overall Reliability score was calculated as the average of these two ratings, with the scale reversed so that higher reliability scores indicated greater appreciation of the importance of reliable measurement.

Experimental control. Judgments about causality necessitate an understanding and evaluation of the conditions under which causality may be validly inferred. Of particular importance is an understanding of the quality of controls exercised in collecting the evidence, which provides the quintessential criterion for the evaluation of the quality of scientific endeavours and the ability to generate informative studies or experiments. To measure understanding of experimental control in this study, students were asked to indicate their confidence in the outcomes of two hypothetical investigations that used different experimental strategies to assess the effects of the two factors. One question described an investigator who isolated the effects of each variable by varying one factor at a time while holding the other constant (the VOTAT strategy) and asked participants to rate their confidence in this experimenter's conclusions. In contrast, a second item asked participants to indicate their confidence in the conclusions of an experimenter who varied both factors at once, never isolating the effects of either factor. For each participant, an Experimental Control score was calculated as the difference between his or her confidence rating for the VOTAT strategy and that for the lesser strategy. Thus, higher scores on this variable indicated greater understanding of the importance of experimental control in evaluating causality.

Objectivity. Judgments about objectivity necessitate an understanding and evaluation of possible conflicts between prior knowledge or biases that one may hold about a phenomenon, or the interests or intentions of others making claims about a phenomenon. Such objectivity may be evaluated with respect to claims made by others, or with respect to one's own beliefs or biases. To measure Objectivity in this study, students were presented the results of a hypothetical experiment and asked to indicate how one's prior beliefs should affect the conclusions drawn from the data. Specifically, two questions were used. The first question asked how relevant, on a scale of 1 to 4 ($1 =$ completely irrelevant and $4 =$ the only relevant factor), the hypothetical experimenter's prior beliefs should be to his conclusions. The second question asked how relevant the participant's own prior beliefs should be to his or her own conclusions about the experiment. An overall Objectivity score was calculated by averaging students' responses to these two questions. These scores were then reversed, so that higher scores represented more advanced understanding of the objectivity of scientific inquiry.

Scientific reasoning. Our assessments of scientific reasoning were based on students' conclusions about the results of a scientifically-sound experiment described in each exercise. For both problems, the 'observed' results indicated that the Belief-Consistent factor (angle of incline for physics, or parents' reading for social science) was *positively* related to the outcome, and that the Belief-Inconsistent factor (ball weight for physics or TV-watching for social science) was *unrelated* to the outcome. Because the effects portrayed for the Belief-Consistent factor were highly consistent with students' prior knowledge, our analyses focused on students' conclusions about the validity of effects observed within the Belief-Inconsistent factor. Students were asked to indicate (1) what the fictional experimenter should conclude about the effect (positive, negative, or no effect) of ball weight or TV-watching on the outcome, and (2) what their own conclusions about the effect (positive, negative, or no effect) of ball weight/TV watching and the outcome would be. Participants' responses to these two (Belief-Inconsistent) questions were classified as correct (no effect) or incorrect (positive or negative effect). Thus, for each problem, two dichotomous measures of scientific reasoning were constructed: the first variable indicated whether the participant identified the correct conclusion for the experimenter given the observed data (Experimenter Conclusion), and the second variable indicated whether the participant drew the correct conclusion him or herself (Personal Conclusion).

Debriefing lecture and quiz

Following completion of the two scientific exercises, the course instructor debriefed the students by delivering a 30-min lecture that outlined both the main points of each exercise as well as the logic behind the experimental designs to which they had been exposed in the two exercises. He covered subjects such as the counter-balanced design, the dependent and independent variables in each exercise, confounding, reliability, and data on their performance on the two tasks. After this explicit instruction about the effects of the variables on the outcomes, a Debriefing Quiz asked students to again indicate their beliefs about the effects of the two variables for each problem.

Reading comprehension
To measure participants' reading abilities, we administered the Nelson-Denny reading test (Brown, Fishco, & Hanna, 1993), a test of vocabulary and reading comprehension skills standardized for students in grades 9 through 16 (college) as well as adults. Scores on the Nelson-Denny have been found to be correlated with performance in this course (Roberts, Suderman, Suderman, & Semb, 1990), thus we wanted to control for the possibility that differences in performance on the two problems were related to students' reading abilities. The Vocabulary subtest contains 80 multiple-choice items, and the Comprehension subtest consists of seven reading passages with 38 multiple choice questions about those passages. The test takes 45 min to complete. The Nelson-Denny has very good psychometric properties, with internal consistency estimates of 0.89 for Vocabulary and 0.81 for Comprehension. Predictive validity is adequate; for example, the Technical Report indicates that moderate to substantial correlations have been found between Vocabulary and Comprehension scores and classroom performance. Vocabulary and Comprehension scores were highly correlated in our sample, thus we used only Vocabulary (raw scores) in the present analyses.

Procedure

The investigation was carried out during four class sessions conducted for this course (Days 1–4). The two science exercises were administered on Days 1 and 2. The presentation of the two science exercises was counter-balanced across subjects and days. On Day 1, students with even course identification numbers completed the physics exercise and students with odd course identification numbers completed the social science exercise. On Day 2, the order was reversed such that even-numbered students did the social science exercise while odd-numbered students did the physics exercise. Students were given 30 min to complete each exercise. No student failed to complete either exercise in the allotted time. On Day 3, the course instructor delivered the 30-min debriefing lecture and then administered the Debriefing Quiz. Finally, on Day 4, the Nelson-Denney reading test (Brown et al., 1993) was administered.

RESULTS

The analyses were carried out in two stages. The first stage provided descriptive information about participants' prior beliefs about the two problems, their understanding of the four methodological concepts, and their scientific reasoning performance. The second stage applied inferential analyses to examine the roles of prior beliefs and the methodological concepts in scientific reasoning.

Prior beliefs

To provide a description of the students' prior beliefs about the two scientific exercises, we calculated the proportions of students that held accurate prior beliefs for each problem, as a function of factor (Belief-Consistent vs Belief-Inconsistent). As expected, almost all students made accurate predictions about the relation between the Belief-Consistent factor and the dependent variable on both problems; 96.2% of the participants thought that balls would roll more quickly down a steep incline than down a shallow one, and 100% believed that children whose parents read to them more frequently would have superior language outcomes to children whose parents read to them less frequently. In contrast, less than one-fourth of the sample held accurate prior beliefs about the Belief-Inconsistent factor for either problem. Nearly 80% of the respondents thought that the weight of the ball would have an effect on how quickly it would roll down an incline (most felt a heavier ball would roll more quickly, although a small proportion indicated that they thought the opposite), and about 75% of the respondents thought that television watching would be related to children's language outcomes (most thought the relationship would be negative, but some thought it would be positive). Thus, these data confirmed our high and low knowledge classifications.

Methodological concepts

Because the four methodological concepts were measured in the context of each problem, mean scores on each concept were computed separately for the Ball problem and the Language problem. These means, as well as zero-order correlations between the measures, are displayed in Table 1. The correlations in Table 1 can be viewed as being composed of two triangular matrices that show the intercorrelations of concepts within each problem, and a square matrix that shows the intercorrelations of the concepts across the two problems. The triangular matrix formed by the intercorrelations of items 1 through 4 show that measures of the methodological concepts were moderately interrelated for the Ball problem, with most correlations falling below 0.30. Essentially the same pattern of results was observed for the concepts within the Language problem (see the triangular matrix formed by the intercorrelations of items 5 through 8). The intercorrelations of concepts across the two problems (shaded area of Table 1) show that cross-problem correlations for individual core concepts were somewhat higher, averaging around 0.40. The means and standard deviations for the measures shown along the bottom of the table indicate that Function of Evidence scores were higher for the Language problem than for the Ball

Table 1. Pairwise correlations between methodological concept measures and mean scores

	1	2	3	4	5	6	7	8
Concepts: Ball								
1 Function of Evidence	1.00							
2 Reliability	−0.04	1.00						
3 Objectivity	0.33***	0.13	1.00					
4 Experimental Control	0.23***	0.16*	0.25***	1.00				
Concepts: Language								
5 Function of Evidence	0.33***	0.22**	0.28***	0.35***	1.00			
6 Reliability	−0.10	0.40***	0.03	−0.05	−0.00	1.00		
7 Objectivity	0.10	0.14*	0.43***	0.14*	0.31***	−0.01	1.00	
8 Experimental Control	0.10	0.21**	0.28***	0.29***	0.28***	−0.03	0.30***	1.00
Mean score (and SD)	1.20 (1.56)	4.63 (1.72)	2.74 (0.79)	2.14 (2.21)	2.00 (1.47)	3.95 (1.55)	2.72 (0.75)	1.02 (1.79)

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.

Table 2. Percentage of students who drew accurate conclusions about effects of the Belief-Inconsistent factor, as a function of problem and question

	Problem	
	Ball	Language
Conclusion of:		
Hypothetical experimenter	86.7	69.2
Own personal belief	66.8	46.0

problem; this difference was statistically significant, $t(421) = -9.36$, $p < 0.0001$. In contrast, both Reliability and Experimental Control scores were higher when assessed in the Ball exercise than in the Language problem, $ts(421) \geq 7.79$, $ps < 0.0001$. Objectivity scores did not differ according to problem.

Scientific reasoning

To examine the students' scientific conclusions in the two exercises, we analysed performance on the two dichotomous measures of scientific reasoning described earlier. The proportions of participants who responded accurately when asked about the experimenter's and their own conclusions (as a function of problem) are shown in Table 2. A majority of the students drew accurate conclusions on the Ball problem, particularly in response to the Experimenter question. Accuracy rates were somewhat lower on the Language problem, especially for the Personal question, with less than half of the participants drawing the appropriate Personal Conclusion. McNemar's tests confirmed that participants more frequently drew accurate conclusions in the Ball problem than in the Language problem, for both the Experimenter and Personal variables, $\chi^2s(1) \geq 21.5$, $ps \leq 0.0001$. Responses to the two exercises, however, were associated. Students who responded correctly to the Experimenter question on the Ball problem were more likely than others to respond correctly to the analogous question on the Language problem, $\chi^2(1) \geq 13.55$, $p \leq 0.001$. Similarly, accurate conclusions on the Personal question on the Ball problem were significantly associated with accurate conclusions on the Personal question of the Language problem, $\chi^2(1) \geq 7.25$, $p \leq 0.01$. More importantly, for both exercises participants' conclusions were more often correct in response to the question asked about the hypothetical experimenter than the question asked about their own personal conclusions, $\chi^2s(1) \geq 29.4$, $ps \leq 0.0001$. Indeed, crossing students' responses to the Experimenter and Personal questions for each problem, 24% of the participants drew correct conclusions for the experimenter but not for themselves on the Ball problem, and 28% did so on the Language problem. These patterns suggest that although most students may have been able to determine the appropriate conclusion in abstract or hypothetical scenarios, some were reluctant to revise their own beliefs about this factor in the face of contradictory evidence.

The relations between scientific reasoning, prior beliefs, and methodological concepts

Inferential analyses examined the extent to which scientific reasoning was predicted by prior beliefs and core scientific concepts. Generalized Estimating Equation (GEE) models (see Liang & Zeger, 1986) were used to examine the effects of prior beliefs and core methodological concepts (between-subjects factors) on the accuracy of the participants'

Table 3. Pairwise correlations between participant measures and methodological concept variables

	Function of Evidence	Reliability	Objectivity	Experimental Control
Ball problem				
Reading ability	0.17*	0.26***	0.22**	0.34***
Year in school	−0.01	0.09	0.10	−0.18**
GPA	0.13	0.09	0.26***	0.21**
Language problem				
Reading ability	0.20***	0.02	0.16**	0.29***
Year in school	−0.03	−0.03	−0.03	−0.07
GPA	0.24***	0.07	0.18***	0.22**

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.

final conclusions. A GEE model is a multivariate or repeated measures form of categorical data analysis, such as logistic regression. The parameter estimates corresponding to the GEE models used here can be interpreted in much the same way as traditional regression coefficients, except that a binomial outcome is being predicted. For example, if one is predicting the likelihood of a correct response, a positive parameter estimate indicates that participants with higher scores on the predictor are more likely to respond correctly than participants with lower scores. GEE models also permit the use of predictors that have varying values across a repeated or within-subject measure. In the current analysis, the methodological concepts were measured separately within each problem, and thus took different values across that within-subjects variable. Separate GEE models were used to predict responses to the Experimenter Conclusion and Personal Conclusion questions.

Preliminary analyses (see Table 3) involved examining the simple correlations between the methodological concept measures and several participant measures, including reading ability (Vocabulary score on the Nelson-Denney), year in school, and GPA. Table 3 shows that GPA and reading ability were consistently correlated with the concept measures. As a result, these measures, as well as problem order (i.e. Day 1 or Day 2) were included as covariates in all analyses. We included gender as an additional covariate because gender differences in science achievement have been well-documented across many cultures (e.g. Friedler & Tamir, 1990; Hassan & Khalifa, 1999; Kahle, Parker, Rennie, & Riley, 1993). Because scientific reasoning performance differed between the Ball and Language exercises, we also controlled for problem in all models. Finally, because the effects of the methodological concepts might vary according to prior beliefs, interactions between the concepts and prior beliefs were included as initial terms in all models tested. However, if they were found to not add significantly to the prediction of a particular outcome variable, they were removed for the final presentation of the analyses.

The results of the GEE model predicting participants' performance on the Experimenter Conclusion variable (i.e. given the data reported, what should the hypothetical experimenter conclude?) are summarized in the left half of Table 4. There was a main effect of reading ability such that participants with better reading skills were more likely to identify the correct conclusion for the hypothetical experimenter than participants with poorer reading abilities. There was also a main effect of problem; students were more likely to identify the correct conclusion for the Ball problem than the Language problem. Two of the methodological concepts predicted performance on this question; higher scores on the

Table 4. Results of GEE analyses predicting Experimenter and Personal Conclusions

	Experimenter Conclusion		Personal Conclusion	
	β	SE β	β	SE β
Control variables				
Gender (Female = 1, Male = 0)	0.56	0.31	0.52	0.28
Reading ability	0.03**	0.01	0.01	0.01
GPA	0.03	0.25	0.21	0.21
Order (Ball-Language = 1)	0.27	0.26	−0.22	0.22
Problem (Ball = 1)	1.15***	0.29	1.16****	0.25
Prior Knowledge (Accurate = 1)	0.23	0.33	3.51**	1.09
Methodological concepts				
Function of Evidence	0.24*	0.10	0.10	0.08
Reliability	0.01	0.08	−0.01	0.07
Objectivity				
Mean level	0.22	0.19	0.86****	0.19
Objectivity × Prior Beliefs (Accurate = 1)[a]	—	—	−0.79*	0.38
Experimental Control	0.30***	0.08	0.05	0.06

****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
[a]This interaction was not significant for the Experimenter Conclusion variable and therefore was not included in the model predicting that outcome.

understanding of Function of Evidence and the understanding of Experimental Control were both associated with a greater likelihood of making the correct conclusions for the hypothetical experimenter. Given that none of the higher-order interactions accounted for variance in this model, they are not included in the final presentation of results here.

The right half of Table 4 shows the results of the GEE model predicting the Personal Conclusion variable (i.e. given the data reported, what should the *you* [the respondent] conclude?). As in the previous model, there was a significant effect of problem, such that students were significantly more likely to draw the correct conclusion for the Ball problem than the Language problem. Unlike the previous model, there was also a significant main effect of Prior Beliefs. Students with accurate prior beliefs about the direction of the factor's effect (i.e. those who believed that the weight of a ball would be unrelated to how quickly it would roll down an incline, or those who believed that television watching would be unrelated to child language outcomes) were more likely to select the appropriate conclusion for themselves than students who had incorrect prior beliefs about these factors. Stated another way, students who held inaccurate prior beliefs were more likely than those with accurate beliefs to draw incorrect Personal Conclusions. Objectivity also predicted performance; students with higher Objectivity scores were more likely to draw the accurate conclusion than students with lower scores. These main effects, however, were qualified by a significant interaction between Prior Beliefs and Objectivity, such that Objectivity was positively associated with drawing accurate personal conclusions only among those students with inaccurate prior beliefs. Thus, the students who were most likely to draw inaccurate Personal Conclusions were those who held inaccurate prior beliefs and scored low on Objectivity.

The presence of a significant effect for the Prior Belief factor for Personal Conclusions, combined with its absence for Experimenter Conclusions, suggested to us that some students were (a) drawing accurate conclusions for what the hypothetical experimenter should conclude from his/her data when those data contradicted the experimenter's prior

Table 5. Number of participants that made accurate and inaccurate conclusions about effects of the Low Knowledge factor on Experimenter Conclusion and Personal Conclusion questions

	Experimenter Conclusion	
	Correct	Incorrect
Ball problem		
Personal Conclusion		
Correct	132	9
Incorrect	51	19
Language problem		
Personal Conclusion		
Correct	86	11
Incorrect	60	54

beliefs about the factor, but (b) did not draw accurate Personal Conclusions in the face of empirical data that contradicted their own prior beliefs. Why would this occur? To address this question, a follow-up analysis was conducted on the subset of individuals who responded correctly to the Experimenter Conclusion variable (i.e. the experimenter should conclude that the weight of the ball is irrelevant to its speed down an incline, or that television watching is unrelated to child language outcomes), to determine what factors distinguished between participants who also gave correct responses to the Personal Conclusion question and those who did not. Table 5 shows the number of participants who correctly and incorrectly responded to these two questions for the Language and Ball problems. We then conducted an analysis on 194 subjects who fell into the left data column of this table. The results of this GEE model on these individuals are shown in Table 6. Once again there was a significant effect of Problem, such that students were more likely to draw the correct conclusions on the Ball problem than on the Language problem. Moreover, there was a strong main effect for Prior Beliefs. This suggests that respondents who identified the correct conclusion for the hypothetical experimenter were also far more likely to draw correct conclusions for themselves if they held a priori beliefs that were

Table 6. Results of GEE analyses predicting Personal Conclusions among participants who responded correctly to the Experimenter Conclusion question

	Personal Conclusion	
	β	SE β
Control variables		
Gender (Female = 1, Male = 0)	0.30	0.33
Vocabulary Score	0.00	0.01
GPA	0.36	0.25
Order (Ball-Language = 1)	−0.27	0.26
Problem (Ball = 1)	0.82**	0.29
Prior Knowledge (Accurate = 1)	1.37****	0.36
Methodological Concepts		
Function of Evidence	0.03	0.09
Reliability	0.03	0.08
Objectivity	0.80****	0.20
Experimental Control	−0.02	0.07

****$p < 0.0001$, **$p < 0.01$.

consistent with the outcome of the experiment. That is, the group of students who identified the correct conclusion for the hypothetical experimenter but who did not identify the correct conclusion for themselves were those whose prior beliefs were contradicted by the outcome of the hypothetical experiment. A significant main effect of Objectivity provided the last interesting finding for this analysis. Students who indicated higher understanding of the concept of objectivity were more likely to draw a Personal Conclusion that was consistent with what they had recommended for the hypothetical experimenter (i.e. that was consistent with the data). Thus, an understanding of the role of objectivity helped participants overcome prior beliefs and maintain consistency in the drawing of accurate personal conclusions.

Performance on debriefing quiz

To determine whether students revised their theories about the causal status of the target variables after explicit instruction, performance on the debriefing quiz was analysed. Twenty-six students for whom we had complete experimental data did not take the debriefing quiz; thus the sample size for this analysis was 185. For the Belief-Consistent factor, 98% of the students responded correctly for the Ball problem and 97% responded correctly for the Language problem. For the Belief-Inconsistent factor, 97% of the participants selected the correct conclusion for both the Ball problem and the Language problem. Thus, in contrast to performance on the scientific reasoning measures, almost all students responded correctly to these questions, even when they were about the causal status of the Belief-Inconsistent factor.

DISCUSSION

This study provides an initial step into the understanding of the relationship between the appreciation of methodological concepts thought to be fundamental to scientific thinking and the quality of scientific reasoning in different contexts. As indicated by the students' responses to the prior beliefs questions, the data bore out our assumptions about the Belief-Consistent and Belief-Inconsistent aspects of the phenomena in question. Almost all students answered the questions about the Belief-Consistent variables correctly, whereas very few responded correctly to questions about the Belief-Inconsistent variables. When provided evidence about the causal status of the Belief-Inconsistent variables, there was considerable variability in college students' abilities to draw accurate, evidence-based conclusions. Although many participants were able to drew accurate scientific conclusions in both the abstract and personal contexts, participants were more likely to reason accurately when asked what a hypothetical experimenter should conclude than when asked what they themselves would conclude. Indeed, on both problems, about one-fourth of the sample made accurate conclusions for the experimenter but not for themselves.

Psychometric data revealed some patterns of interest with respect to the measurement of the methodological concepts that are generally thought to form a basis for scientific reasoning. Zero order correlations of the derived measures of Function of Evidence, Objectivity, Reliability, and Experimental Control were modestly but significantly correlated with one another within problems, with measures of the understanding of Function of Evidence, Objectivity, and Experimental Control tending to hang together in each of the dimensions as an identifiable cluster of measures. Furthermore, concepts measured across problems (e.g. Function of Evidence across the Ball and Language problems) were moderately correlated with one another, suggesting that each measure tapped a convergent concept that transcended the problems themselves. Finally, in general, performance on the methodological concepts was modestly but significantly related to general academic achievement variables, such as reading ability and GPA.

The most interesting findings concerned the degree to which understanding of these concepts could be brought to bear on the quality of scientific reasoning. Here, we observed a pattern of results that suggested that there was some dissociation between what predicted accurate performance in students' drawing of accurate conclusions about phenomena *in the abstract* versus the drawing of accurate conclusions *for themselves*, especially in cases where the data contradicted existing prior beliefs about those phenomena. Essentially, the results indicate that two of the methodological components (i.e. understanding the function of evidence and the concept of experimental control) were significant predictors of students' recommendations for what a hypothetical experimenter should conclude about a phenomenon in the face of data that (s)he collected, even if those data contradicted the experimenter's existing prior beliefs about the phenomenon in question. As such, the findings appear to be encouraging with respect to the development of a general model of scientific reasoning based on these underlying methodological concepts.

Students' reliance on these concepts, however, seemed to change when asked about what *they themselves should conclude* about a phenomenon from scientifically collected data that conflicted with their prior beliefs. The students were less likely to draw the correct conclusion when asked about their own conclusions than when asked about the hypothetical experimenter, and accuracy in personal conclusions was predicted by a combination of their own prior beliefs (correct prior beliefs were associated with the drawing of valid personal conclusions from the experiments described), and the extent to which they understood objectivity of scientific inquiry. Specifically, among those students with inaccurate prior beliefs, those with an appreciation of objectivity were better able to distance themselves from their beliefs and draw accurate conclusions than those with a lesser understanding of objectivity. Thus, consistent with much previous research, many college students were reluctant to revise their preexisting naïve theories about phenomena, even in the face of contradictory evidence (e.g. Klahr & Dunbar, 1988; Kuhn et al., 1988; Schauble, 1996; Winer et al., 2002).

Taken together, these findings suggest that many students understood, in the abstract, how data should relate to the validity of scientific conclusions about phenomena, but when applied to their own conclusions, their prior beliefs about those phenomena typically took precedence over the description of data or findings (however scientifically valid) concerning those phenomena, especially when they did not understand the importance of objectivity. This conclusion is further bolstered by an analysis of students who identified the valid/correct conclusion for the hypothetical experimenter but who drew invalid/incorrect conclusions for themselves. Those students were more likely than others to have prior beliefs about the phenomenon that were in conflict with the data described and to have a poorer understanding of objectivity.

Why might students have been more biased by their prior beliefs when drawing their own conclusions than when recommending a conclusion for the experimenter? Several possibilities can be discussed.

First, the observed differences in reasoning in abstract and personal contexts might reflect students' reliance on two different cognitive frameworks. For example, two-process theories of scientific reasoning assume that a heuristic reasoning system that relies on intuitions, 'rules of thumb,' and available memories coexists with a more accurate analytic reasoning system (e.g. Kahneman, Slovic, & Tversky, 1982; Klaczynski, 2000, 2001; Stanovich, 1999; Stanovich & West, 1998). According to these models, because decontextualized thinking requires analytic processing, the interference of prior beliefs in accurate reasoning is the result of a reliance on the heuristic, rather than the analytic, system. A variety of factors, such as individual difference variables, processing load, social demands for accuracy, and the extent to which a problem activates personal theories are all thought to determine which system predominates (e.g. Klaczynski, 2000, 2001). For example, in the current study the personal context might have been more likely than the abstract context to activate personal theories and reliance on the heuristic system.

A second possibility comes from the work of Kuhn and her colleagues (Kuhn et al., 1995), which suggests that high levels of affective investment in one's theories can lead to difficulty in metacognitive distancing. Although this explanation makes no claims about the existence of two processing systems, it is not necessarily inconsistent with two-process models. Thus, students might have had greater affective involvement in their own conclusions than in those of a hypothetical experimenter, leading them to have greater difficulty distancing themselves from their prior beliefs when asked what they concluded than when asked what the experimenter should conclude about his data. Alternatively, students' responses to the experimenter question might have been the result of demand characteristics whereas their responses to the personal question were based on what they *really* believed about the phenomena.

Finally, it is also possible that students who reasoned accurately for the experimenter question but not for the personal question simply viewed the experimenter's results as truly hypothetical; that is, these students might not have treated the 'observed results' as real scientific data worthy of prompting personal theory revision. The fact that students' grasp of objectivity also predicted accurate personal conclusions, however, argues against this last interpretation.

Consistent with the literature on reasoning and decision-making, the current study suggests that college students' abilities to reason independently of prior beliefs vary as a function of both problem context and individual difference factors. The role of objectivity in predicting resistance to prior belief is especially interesting in light of other research on individual differences in reasoning biases. Several investigators (e.g. Klaczynski, 2000; Perkins et al., 1993; Stanovich & West, 1998) have proposed that unbiased reasoning is associated with a thinking style or metacognitive disposition that involves reflectiveness, open-mindedness, and a general tendency to regard knowledge-acquisition as more important than belief preservation. Consistent with this hypothesis, Stanovich and West (1998) found that a measure of the disposition to engage in 'actively open-minded thinking,' was associated with performance on a battery of reasoning and decision making tests. Thus, it is possible that our measure of the understanding of objectivity taps a personal disposition towards open-mindedness. This issue is worthy of additional research, as it has implications for the amenability of objectivity to instruction. The reasoning literature also suggests other factors that might be related to decontextualized or unbiased reasoning, including general cognitive ability and the extent to which problems activate highly personal beliefs (Klaczynski et al., 1997; Kuhn et al., 1995; Sa et al., 1999; Stanovich & West, 1997, 1998). For example, this work suggests that we might have observed lower rates of unbiased reasoning if the scientific exercises had activated prior beliefs that were of high personal relevance to the participants, such as those associated with political, social, or religious issues.

Interestingly, overall accuracy rates on both the Experimenter Conclusion and Personal Conclusion question were higher for the Ball problem than for the Language problem. Because we did not systematically isolate the features that might distinguish between these two problems, the current data provide no clear explanations for this pattern. It is possible that the results reflect more general domain (i.e. physics vs social science) differences in scientific reasoning. Consistent with this interpretation, Kuhn and her colleagues (1995), presented adults and children scientific reasoning problems in a physical domain (e.g. determining the factors affecting the speed of a boat through water) and in a social domain (e.g. determining the factors that affect children's ratings of television programmes) and found that both children and adults performed more poorly in the social domain. Alternatively, the problem effects observed here might reflect differences in participants' beliefs about the importance of data and the quality of data gathered in the two problems. For instance, the hypothetical research described in the Language problem was correlational in nature, whereas the research described in the Ball problem was a true experiment. This interpretation is reinforced by the problem differences in scores on the methodological concepts themselves. Finally, it is quite possible that the differences in performance were simply the result of differences in the students' comprehension of the two problems.

Following explicit instruction about the causal status of variables in the debriefing lecture, almost all students selected the correct conclusions in the debriefing quiz. Of course, we do not know whether students correct responses were the result of revisions to their own theories about the target variables (i.e. true belief change) or the incentive for providing the correct answer (i.e. a good grade on the quiz). In either case, from an educational perspective, students' correct responses to questions about the causal status of variables in scientific problems such as these represent some progress, as instantiating these concepts in the knowledge base may, in turn, become part of the prior knowledge they bring to similar problems in the future, both real and hypothetical.

These data provide interesting, if not somewhat disconcerting, findings with regard to the dissociation of the quality of scientific reasoning in the abstract and the drawing of accurate personal conclusions in the face of one's own prior beliefs about a particular phenomenon. The findings have interesting implications that apply across a wide range of issues related to science education. For example, the data here suggest that it is entirely possible to engage students in a curriculum on scientific reasoning and have them perform at a high level on the general concepts when tested about those concepts and the generation of scientifically valid conclusions, yet fail to use those concepts effectively when reasoning scientifically about real-world issues—particularly those with scientific solutions that may be at odds with their own existing prior beliefs. It remains to be seen whether a hands-on, laboratory course would be more effective in promoting students' use of those concepts in their own lives, but the data here suggest strongly that even if a laboratory course were to be implemented for the conveyance of these concepts, it would be important for students to conduct a study or experiment in which the results were clearly at odds with their own expectations.

ACKNOWLEDGEMENTS

The authors wish to thank Janis Bulgren and James Ellis of the School of Education at the University of Kansas for helping to initiate this collaboration. In addition, we thank four anonymous reviewers for their constructive comments on earlier versions of this manuscript.

REFERENCES

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brown, J. I., Fishco, V. V., & Hanna, G. S. (1993). *The Nelson-Denny Reading Test Forms G and H.* Chicago, IL: Riverside Publishing Co.

Byrnes, J. P. (2001). *Cognitive development and learning in instructional contexts*. Boston: Allyn and Bacon.

Chen, Z., & Klahr, D. (1999). All other things being equal: acquisition and transfer of the Control of Variables Strategy. *Child Development, 70*, 1098–1120.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1–49.

Friedler, Y., & Tamir, P. (1990). Sex differences in science education in Israel: an analysis of 15 years of research. *Research in Science and Technological Education, 8*, 21–34.

Hassan, M. M., & Khalifa, A. (1999). Sex differences in science achievement across ten academic years among high school students in the United Arab Emirates. *Psychological Reports, 84*, 747–757.

Kahle, J. B., Parker, L. H., Rennie, L. J., & Riley, D. (1993). Gender differences in science education: building a model. *Educational Psychologist, 28*, 379–404.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertaintly: Heuristics and biases*. Cambridge, MA: Cambrige University Press.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: a two-process approach to adolescent cognition. *Child-Development, 71*, 1347–1366.

Klaczynski, P. A. (2001). Analytic and heuristic processing influences on adolescent reasoning and decision-making. *Child-Development, 72*, 844–861.

Klaczynski, P. A., Gordon, P. A., & Fauth, J. (1997). Goal-oriented critical reasoning and individual differences in critical reasoning biases. *Journal of Educational Psychology, 89*, 470–485.

Klahr, D., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science, 12*, 1–48.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: a developmental study. *Cognitive Psychology, 25*, 111–146.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: the role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development, 60*, 1316–1327.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674–689.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* New York: Academic Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4).

Leshowitz, B., DiCerbo, K. E., Okun, M. A. (2002). Effects of instruction in methodological reasoning in information evaluation. *Teaching of Psychology, 29*, 5–10.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: a study with sinking objects. *Child Development, 67*, 2709–2727.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: a dispositional theory of thinking. *Merrill-Palmer Quarterly, 39*, 1–21.

Rice, M. (1983). The role of television in language acquisition. *Developmental Review, 3*, 211–224.

Roberts, M. S., Suderman, L., Suderman, R., & Semb, G. B. (1990). Reading ability as a performance predictor in a behaviorally based psychology class. *Teaching of Psychology, 17*, 173–175.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341–357.

Sa, W. C., West, R. F., & Stanovich, K. E. (1999). The domain-specificity and generality of belief bias: searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497–510.

Schauble, L. (1990). Belief revision in children: the role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31–57.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102–119.

Shannon, P., & Fernie, D. E. (1985). Print and television: children's use of the medium is the message. *Elementary School Journal, 85*, 663–672.

Sigel, I. (1993). The centrality of a distancing model for the development of representational competence. In R. Cocking, & K. A. Renninger (Eds.), *The development and meaning of psychological distance*. Hillsdale, NJ: Erlbaum.

Singer, J. L., & Singer, D. G. (1983). Psychologists look at television: cognitive, developmental, personality, and social policy implications. *American Psychologist, 38*, 826–834.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychological Review, 111*, 42–61.

Stanovich, K. E. (1999). *Who is rational: Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342–357.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161–188.

Wells, G. (1985a) *Language development in the preschool years*. Cambridge, UK: Cambridge University Press.

Wells, G. (1985b). Preschool literacy-related activities and success in school. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing* (pp. 229–255). Cambridge, UK: Cambridge University Press.

Winer, G. A., Cottrell, J. E., Gregg, V., Fournier, J. S., & Bica, L. A. (2002). Fundamentally misunderstanding visual perception: adults' beliefs in visual emissions. *American Psychologist, 57*, 417–424.

Wright, J. C., Huston, A. C., Murphy, K. C., St. Peters, M., Pinon, M., Scantlin, R., & Kotler, J. (2001). The relations of early television viewing to school readiness and vocabulary of children from low-income families: The early window project. *Child Development, 72*, 1347–1366.

