1-1-2011

Examining the Relationship Between Fact Learning and Higher Order Learning via Retrieval Practice

Pooja Agarwal
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Examination Committee:
Henry L. Roediger, III (Chair)
David A. Balota
Susan M. Fitzpatrick
Mark A. McDaniel
R. Keith Sawyer
Michael J. Strube

EXAMINING THE RELATIONSHIP BETWEEN FACT LEARNING AND HIGHER ORDER LEARNING VIA RETRIEVAL PRACTICE

by

Pooja Kay Agarwal

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2011

Saint Louis, Missouri

Abstract

The development of higher order skills is a desired outcome of education. Some believe that higher order learning can be improved directly, whereas others argue that higher order learning can be improved via the enhancement of factual or conceptual knowledge. The relationship between fact and higher order learning is often speculated, but empirically unknown.

This project examines whether retrieval practice via quizzing, a strategy typically used to enhance fact learning, can be used as a strategy to improve higher order skills in both laboratory and applied settings. In the current study, higher order skills were considered to comprise the *understand*, *apply*, *analyze*, *evaluate*, and *create* categories of a revised Bloom's taxonomy of knowledge and skills in education (Anderson et al., 2001). Across three experiments, subjects engaged in retrieval practice with basic concept questions, higher order questions, or a mix of question types. Performance was measured after a two-day delay on both concept and higher order questions in order to determine the type of retrieval practice that produced the greatest level of delayed performance.

Retrieval practice (regardless of question type) improved both delayed concept and higher order test performance more than restudying or no quizzing. In Experiments 1 and 2 with college students, delayed performance was greatest when the initial quiz question type matched the final test question type, consistent with a pattern of transfer-appropriate processing; however, benefits from conceptual retrieval practice on delayed higher order performance or from higher order retrieval practice on delayed concept performance were not found. In

Experiment 3 with middle school students, a mix of concept and higher order quiz questions produced the greatest long-term learning, although this improvement was only marginally greater than the benefit from higher order retrieval practice on delayed higher order learning.

The current project is the first to demonstrate that retrieval practice with higher order questions improves delayed performance on higher order test questions from complex taxonomic categories. Retrieval practice can be used as a strategy to enhance both conceptual and higher order skill learning, and teachers are encouraged to apply retrieval practice strategies in their classrooms.

Acknowledgements

First, I thank my primary advisor and committee chair, Roddy Roediger, for his guidance and support throughout the years, as well as my dissertation committee members for their helpful input: Dave Balota, Mark McDaniel, Mike Strube, Susan Fitzpatrick, and Keith Sawyer. I am grateful to Columbia Middle School teacher Patrice Bain, principal Roger Chamberlain, and the 2010-2011 6th grade students for their patience and participation. I also thank the Roediger Memory Lab and the Balota Cognitive Psychology Lab for valuable discussions and assistance with data analyses. Finally, I thank my family and friends for their love and encouragement. This work is dedicated in memory of my father, Ashok K. Agarwal, M.D., and also in honor of my grandfather, Harish C. Mital, Ph.D.

This research was supported by the Harry S. Truman Scholarship Foundation, the National Science Foundation Graduate Research Fellowship Program, and the James S. McDonnell Foundation 21st Century Science Initiative grant, Applying Cognitive Psychology to Enhance Educational Practice: Bridging Brain, Mind, and Behavior Collaborative Activity Award.

Please address correspondence to Pooja K. Agarwal, Washington University in St. Louis, One Brookings Drive, Campus Box 1125, St. Louis MO, 63130, (314) 443-7848, pooja.agarwal@wustl.edu, http://www.poojaagarwal.com.

Table of Contents

List of Tables

List of Figures

List of Appendices

Examining the Relationship Between Fact Learning and Higher Order Learning via Retrieval Practice

Recent research has demonstrated the robust effects of retrieval practice for enhancing long-term learning (McDaniel, Roediger, & McDermott, 2007; Roediger, Agarwal, Kang, & Marsh, 2010; Roediger & Karpicke, 2006b; Rohrer & Pashler, 2010). Typically, students study a set of material (e.g., word pairs, foreign language vocabulary words, prose passages), engage in retrieval practice via quizzing or testing, and then immediately or after a delay (ranging from hours, to days, to weeks), students complete a final criterial test. In general, retrieval practice improves final performance and increases retention for a variety of student populations, materials, and time delays when compared to restudying or no quizzing (see Roediger & Karpicke, 2006a, for a review).

A recurring criticism from educational researchers and practitioners is that retrieval practice only enhances knowledge or memory of the to-be-studied material (e.g., Gatto, 2011). Is it the case, however, that retrieval practice can also promote higher order skills such as problem solving and critical thinking? In addition, can retrieval practice benefit higher order skills indirectly, perhaps by first improving factual knowledge? That is, students with a larger knowledge base of readily accessible facts and concepts (strengthened by retrieval practice) may be more proficient at demonstrating higher order skills than students with a less developed knowledge base. The purpose of this study is to examine whether retrieval practice can be used as a technique to improve higher order

skills, and also to examine the relationship between fact learning and higher order skill learning.

First, I discuss some current definitions for “higher order skills” and how they can be operationally defined. Next, I provide an overview of the current debate between focusing classroom instruction on fact learning versus focusing instruction on higher order skill learning. Then, I highlight recent findings from the retrieval practice literature relevant to higher order skill improvement; and finally, I provide an overview of two theories, transfer appropriate processing and cognitive load theory, which were used as frameworks for understanding how fact learning may subsequently enhance higher order skill performance. Note that throughout the study, I use the terms “fact” and “concept” interchangeably to refer to “basic” learning (learning in the *remember* and *understand* categories of the revised Bloom’s taxonomy, discussed next).

Higher Order Skills

While there are few agreed-upon definitions of higher order skills (see Agarwal, 2011, for a review), higher order skills are frequently classified using *The Taxonomy of Educational Objectives* by Bloom, Engelhart, Furst, Hill, and Krathwohl (1956). The original taxonomy included six categories of cognitive skills, ranging from simple to complex: *knowledge*, *comprehension*, *application*, *analysis*, *synthesis*, and *evaluation*. Bloom et al. explained that the taxonomy was designed as a step process; to achieve a higher objective or category, one must first master the skills at a lower category. In other words, before comprehension, application, or analysis can take place, a student must first

acquire knowledge. The current study is designed to directly examine Bloom et al.'s argument that an understanding of facts is required in order to perform cognitive skills at complex levels.

In 2001, Anderson et al., including co-authors of the original taxonomy, proposed a revised taxonomy of educational objectives (see Figure 1). The simplified taxonomy highlights educational skills in verb tense: *remember*, *understand* (previously called comprehension), *apply*, *analyze*, *evaluate*, and *create* (previously called synthesis and reordered with evaluation). Within Anderson et al.'s revised taxonomy, higher order skills are considered to comprise the *apply*, *analyze*, *evaluate*, and *create* categories. On the other hand, the skills of recognizing, remembering, and comprehending information fall under Anderson et al.'s *remember* and *understand* categories, skills which are typically measured in retrieval practice research (discussed in the next section).

To examine higher order skill learning, long passages used in middle school and college classrooms were used. Higher order test questions were developed in accordance with the *apply*, *analyze*, *evaluate*, and *create* categories of Anderson et al.'s (2001) revised taxonomy (see Figure 1 for specific skills required of each category). Questions classified within these four categories were operationally defined (see pp. 26-29) and used to engage a spectrum of higher order skills. By including questions from higher categories in Anderson et al.'s (2001) revised taxonomy, the current study was designed to extend our current understanding of retrieval practice and its potential to enhance both fact learning and higher order skill learning.

Thinking with the Basics versus Thinking is Basic

Given the current climate surrounding the ubiquitous use of standardized testing, it is no surprise that many educators would like to move away from "teaching to the test" and "drill and kill strategies" aimed at enhancing basic knowledge learning (Kohn, 1999; Kuhn, 2005). Instead, educators advocate for a shift to classroom time focused on developing higher order skills and a shift toward a variety of assessment techniques, such as essays, projects, papers, and ongoing assessments, to measure students' higher order skills. At the same time, cognitive psychologists often argue that in order to foster higher order skills, we must focus on and reinforce basic knowledge and fact learning (Willingham, 2009, Chapter 5). To spend increased classroom time on higher order skills to the detriment of fact learning, some argue, would be a disservice to our students. The debate between a focus on higher order skills vs. fact learning is often framed in all-or-nothing terms. Greeno (1992) summarized this debate well, and he is worth quoting at length:

> There is widespread agreement that students do less scientific and mathematical thinking than we wish they would. There are, however, two quite different views about the relation of thinking to classroom learning in mathematics and science, which I will call "Thinking with the basics" and "Thinking is basic." According to "Thinking with the basics," the job of classroom learning is to provide basic scientific or mathematical knowledge that students can then use in thinking mathematically or scientifically after they have learned enough and if they are sufficiently

> talented and motivated. According to “Thinking is basic,” learning to think scientifically and mathematically should be a major focus of classroom activity from the beginning… The two views differ in assumptions that they presuppose about the relation of knowledge and thinking. “Thinking with the basics” presupposes quite a sharp distinction between knowledge and thinking, with the possibility of acquiring a great deal of knowledge without much ability [or necessity] to think, but not conversely… “Thinking is basic,” on the other hand, considers ability to think as a natural human endowment, along with other abilities such as locomotion and communication. (pp. 39-40)

To summarize, one side (often cognitive psychologists) argues that it is important for children to have strong knowledge of the basics in order to support higher order skills and thus we should focus classroom time on teaching the basics, while the other side (often educators) argues that it is important to develop children’s critical thinking skills and understanding, and thus we should focus on providing classroom opportunities that directly encourage these skills. Even so, upon closer inspection of these disparate literatures (provided in the next sections), rarely is anyone advocating for 100% teaching of facts or 100% teaching of thinking. Instead, both sides of the debate often meet in the middle and advocate for direct instruction of facts *and* of higher order skills, while in the context of meaningful situations and domain knowledge.

“Thinking with the Basics” Viewpoint

The debate about how heavily our education system and teaching practices should focus on higher order skills vs. basic knowledge may have begun more than 100 years ago. William James (1900) observed,

> The excesses of old-fashioned verbal memorizing, and the immense advantages of object-teaching in the earlier stages of culture, have perhaps led those who philosophize about teaching to an unduly strong reaction; and learning things by heart is now probably somewhat too much despised. (p. 131)

James went on to explain the importance of memorizing to enable students to articulate facts, quotations, and formulas when needed, but he also observed that connecting these facts with other information serves an important basis for critical thinking. Similarly, Hugo Münsterberg (1909) reflected that facts should not be taught in isolation, but instead, students should be encouraged to reflect on the relationships and connections between facts to aid understanding of a topic. Edward Thorndike (1906) also observed that fact learning is only a small part of education; good teachers, instead, focus on teaching and assessing comprehension and understanding, as well as fact learning, but that this blend of teaching (of both facts and understanding) “is one of the hardest things to do well” (p. 260; cf. Glaser, 1984, for Thorndike’s possible influence on drill methods via behaviorism). In addition, Bloom et al. (1956) argued for the importance of higher order skill development, but they also argued that knowledge is a prerequisite for higher order skills, such as comprehension, application, analysis,

etc. For these prominent psychologists (see also Ausubel, 1961/1965; Bartlett, 1958; Bruner, 1959/1965, 1977; Hirsch, 1996; Willingham, 2009), fact learning is simply one important component of a comprehensive education system that should also include comprehension, understanding, and critical thinking.

How, more specifically, does factual knowledge lead to understanding? Willingham (2009) distinguished rote knowledge from the type of knowledge required for higher order skills (see also Ausubel, 1961/1965; Ausubel, Novak, & Hanesian, 1978; Mayer, 2002). According to Willingham, for instance, rote knowledge of facts is simple, isolated, and does not lead to greater understanding of a topic. For instance, knowing that George Washington was the first president of the United States does not lead to a deeper understanding of United States civics or government. Willingham stated that this type of rote fact learning is not the kind that he or others encourage when advocating for fact learning. Instead, what Willingham called shallow knowledge (e.g., presidents are leaders who make important decisions) is what builds on deep knowledge (e.g., if George Washington was the first president, he must have made many important decisions) in order to construct rich understanding of a topic. This deep knowledge is the kind that teachers would like to impart to their students because it presumably provides a foundation for higher order skills such as abstraction, application, and inferencing. In other words, simply learning rote facts or knowledge without connecting them to a deeper knowledge structure may not benefit students' understanding.

In addition, Willingham (2009) argued that when students practice facts until they are memorized, students can more easily apply deeper knowledge to higher order skills (see also Glaser, 1984). Practice makes retrieval of facts (e.g., George Washington was the first president of the United States) and procedures (e.g., using the distributive property in algebra) automatic, thereby requiring less effort and capacity from working memory, and enabling the student to use the additional working memory capacity for more complex skills. In sum, Willingham agreed that students need to develop higher order skills, and he argued that in order to do so, students need basic knowledge of a topic and need to practice retrieving this knowledge in order to focus their cognitive resources on higher order skills. Similarly, as Sternberg, Grigorenko, and Zhang observed (2008),

> Teachers need to move beyond the false dichotomy between "teaching for thinking" and "teaching for the facts"... One cannot analyze what one knows if one knows nothing. One cannot creatively go beyond the existing boundaries of knowledge if one cannot identify those boundaries. And one cannot apply what one knows in a practical manner if one does not know anything to apply. (p. 487)

Thus, according to many psychologists, learning facts in isolation does not benefit higher order skills, and teaching students higher order skills without an existing foundation of knowledge does not produce successful skills, either.

This brief section served to summarize the "thinking with the basics" viewpoint that knowledge of facts is essential and precedes successful execution

of higher order skills. Although the "thinking with the basics" viewpoint is often characterized as supporting strictly fact learning, the current review suggests that these advocates do not adopt such a strict perspective. In general, the arguments advanced by this group of researchers emphasize that fact learning is important in that it underlies higher order skills, particularly when shallow knowledge is connected to deeper knowledge. Next, I consider the alternative viewpoint that thinking is basic and that education should include direct instruction on higher order skills.

"Thinking is Basic" Viewpoint

Proponents of the "thinking is basic" viewpoint hold that substantially less time should be spent on fact learning than is currently the norm, and instead, more time should be afforded to classroom activities that promote thinking, analyzing, and metacognition (for a history of this movement, see Cuban, 1984). Instead of focusing on knowledge learning, John Dewey (1916/1944) recommended, "The sole direct path to enduring improvement in the methods of instruction and learning consists in centering upon the conditions which exact, promote, and test thinking. Thinking *is* the method of intelligent learning, of learning that employs and rewards mind" (p. 153).

Alfie Kohn (1999), a prominent figure in the area of progressive and constructivist education, also supports the "thinking is basic" viewpoint and he argued that the ability to reference information and use it is much more valuable than the isolated facts themselves. At the same time, he maintained that it is important to ascertain what the "basics" are; even if teachers focus their teaching

on thinking, students must have something to think *about*. Again, Kohn's argument is not whether we should teach facts at all; rather, it's how much time should be spent on them.

Because time spent on both fact learning and thinking is valuable, Kohn (1999) recommended that teachers should spend time on fact learning, but in the context of answering meaningful, engaging questions that require students' higher order skills, instead of questions that ask for knowledge of an isolated fact. Kohn used the familiar example of dividing fractions; often, children (and adults) have learned the procedure for dividing fractions by multiplying by the reciprocal, but they lack the understanding to explain why this procedure works and why you get a larger fraction after division. In this example, students have memorized a simple fact or procedure, but lack understanding. In general, Kohn fits firmly within the "thinking is basics" camp, and while he advocates for a decrease in classroom time spent on memorizing basic facts, his writings suggest that a balance between fact learning in the context of higher order skills and complex situations would be ideal for student learning.

In sum, the "debate" is not whether to teach facts in isolation or to teach thinking in isolation; instead, there appears to be some extant debate on how much classroom time to spend on fact learning vs. how much time to spend higher order skills, but all agree that facts should be taught in meaningful contexts, and that, to paraphrase Kohn (1999) and Sternberg et al. (2008), in order to think, we must have something to think *about*.

Retrieval Practice

Test-enhanced learning, or the use of tests and quizzes to engage retrieval processes, has been widely demonstrated as an effective strategy for facilitating fact learning (McDaniel, Roediger, & McDermott, 2007; Roediger et al., 2010; Roediger & Karpicke, 2006a), but research using this strategy has been largely limited to demonstrating benefits of retrieval practice on retention and not on higher order skills. If retrieval practice via quizzing enhances long-term retention of knowledge compared to no testing or even restudying (typically referred to as the testing effect), retrieval practice may also benefit higher order skills.

Recently, a few studies have examined the effect of retrieval practice on cognitive skills in taxonomic categories other than the *remember* category from Anderson et al.'s (2001) revised Bloom's taxonomy. For instance, Butler (2010) conducted three experiments that demonstrated positive effects of retrieval practice on both fact learning and far transfer to a new knowledge domain (i.e., the *apply* taxonomic category of Anderson et al.'s, 2001, revised taxonomy). In Experiments 1 and 2, subjects were asked to read a set of six passages (approximately 1,000 words in length) about various topics, followed by restudying half of the passages three times or completing three short answer tests (which included both factual and conceptual questions, followed by feedback) on the other passages. On a final test after one week, repeated testing led to better retention of facts and concepts, and better transfer to new

inferential questions within the same knowledge domain, than repeated studying, with differences ranging between 20-50%.

In Experiment 3, Butler (2010) extended his findings by including final inferential test questions from a different knowledge domain (e.g., a initial test question about the wing structure of bats versus birds, a final test question about the U.S. Military designing new aircraft wings based on bats versus jet fighters). Again, repeated testing led to better transfer performance on inferential questions than repeated studying (a difference of 24%), even when final test questions were from a novel knowledge domain; this latter experiment demonstrates far transfer along the knowledge domain (see Barnett & Ceci, 2002). In addition, according to Butler, perhaps by using retrieval practice to increase the retention of factual and conceptual information, the process of transfer (applying information to novel situations) became easier for subjects to execute.

In another recent study examining the effect of retrieval practice on more complex cognitive skills, Jacoby, Wahlheim, and Coane (2010) demonstrated positive effects of retrieval practice on subjects' classification skills (i.e., one component in the *understand* category of Anderson et al.'s, 2001, revised taxonomy). In Experiment 1, college students studied 80 bird exemplar picture-family name pairs (e.g., 10 exemplar bird pictures were presented for the *thrasher* family of birds), in either a repeated study condition or a repeated testing condition. In the repeated testing condition, subjects were provided with a picture of a bird and its family name on the first trial, and were instructed to recall the family name on subsequent test trials (followed by correct/incorrect feedback

with the family name). On an immediate final test, subjects made recognition memory decisions (old/new) and classification decisions for studied and novel exemplars, where subjects were presented with all eight family names used in the experiment and were asked to select the family to which the exemplar belonged.

In the Jacoby et al. study (2010), final recognition performance of studied exemplars was greater following repeated testing (81%) than restudying (72%). In addition, classification of both studied and novel exemplars was significantly greater following the repeated testing condition (79% and 53%, respectively) than the restudying condition (71% and 45%, respectively). In Experiment 2, Jacoby, et al. replicated the results from Experiment 1, and additional testing (five tests instead of three tests) further enhanced the benefits of retrieval practice for recognition memory and classification performance. Jacoby et al. concluded that even with more complex materials (i.e., natural concepts like bird families) than those typically used in testing effect experiments, retrieval practice still enhanced recognition memory and higher order classification skills more than restudying.

This brief review suggests that retrieval practice can be used to enhance fact learning, transfer to new contexts (Butler, 2010; see also Rohrer, Taylor, & Sholar, 2010), and classification skills (Jacoby et al., 2010). A number of other strategies used to enhance fact learning, for instance spaced practice (Rohrer & Taylor, 2006) and interleaved practice (items studied repeatedly with intervening items; Kornell & Bjork, 2008; Taylor & Rohrer, 2010) have also been used to enhance higher order processes that fall under Anderson et al.'s (2001)

understand category (see Agarwal, 2011, for a review). Whether retrieval practice can benefit additional higher order skills, however, remains an open question. Thus, the present experiments include test questions from more complex categories in Anderson et al.'s taxonomy (*apply*, *analyze, evaluate,* and *create*), designed to engage higher order skills such as differentiating, critiquing, and predicting, skills that have not been examined within a retrieval practice paradigm to date.

It is important to note that in the two studies outlined above, as well as others (e.g., Chan, McDermott, & Roediger, 2006; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2011), initial quiz items were "yoked" or related to final transfer test items in order to directly examine the effect of retrieval practice using one item on the learning of an associated but different item. Complex higher order skills, on the other hand, require synthesis across a number of key ideas in order to evaluate, analyze, make predictions, etc. Thus, in a point of departure from past research, the present experiments included initial quiz questions that were comprised of key ideas or concepts stated explicitly in the passages (i.e., as opposed to minute factual details, such as names, dates, definitions, etc.; hereby referred to as "concept questions"), as well as final higher order test questions that were broad and required subjects to draw on a number of key ideas or concepts in order to respond (hereby referred to as "higher order questions").

In other words, items in the present experiments were not yoked from one concept item to one higher order item; instead, many concepts comprised the

ideas required to answer a higher order item, in order to mimic realistic educational materials. In addition, while concept questions were used in the current study to encourage deep knowledge and meaningful learning, in accordance with Willingham (2009) and Kohn (1999) discussed earlier, the terms "fact" and "concept" are used interchangeably to refer to the relationship between basic learning, in contrast to higher order learning.

Given the above consideration of higher order skills, as well as retrieval practice as a potent memory modifier, I now turn to the theoretical rationale for understanding how retrieval practice may serve as a potential strategy for enhancing higher order skills.

Theoretical Rationale

In the present three experiments, subjects engaged in retrieval practice with only concept questions or only higher order questions (Experiment 1), or a mix of question types (in Experiments 2 and 3). Performance was measured after a two-day delay on both concept and higher order questions in order to determine the type of retrieval practice that produced the largest gain in delayed performance. Two theories were used as frameworks for understanding how fact and concept learning (via retrieval practice) may subsequently enhance higher order skills: transfer appropriate processing and cognitive load theory.

Transfer Appropriate Processing

The transfer appropriate processing framework states that final performance should be optimized when encoding processes engaged during learning match retrieval processes engaged during testing (Morris, Bransford, &

Franks, 1977; see also McDaniel, Friedman, & Bourne, 1978). This framework is often cited as an explanation for why retrieval practice enhances long-term retention – by engaging in retrieval practice during study, students can match their initial processing to the type of processing required at test (Roediger & Karpicke, 2006a). Using transfer appropriate processing as a framework for the current study, retrieval practice with concept questions should benefit delayed concept performance, and retrieval practice with higher order questions should benefit delayed higher order performance (compared to no quizzing or restudying in both cases). Note that the latter situation (higher order retrieval practice may enhance delayed higher order test performance) supports the "thinking is basic" viewpoint, discussed earlier.

In addition, the transfer appropriate processing framework suggests that engaging in higher order retrieval practice may enhance both concept and higher order performance on the delayed test. Roediger and Karpicke (2006a) explained that retrieval practice with recall tests generally promotes performance on both recognition (e.g., multiple-choice) and recall tests. This finding is not necessarily counter to transfer appropriate processing, if retrieval practice with recall tests engages processes required for both recognition and recall tests, such as retrieval effort, to a greater extent than does retrieval practice with recognition tests.

Similarly, both the retrieval effort hypothesis (Gardiner, Craik, & Bleasdale, 1973; Pyc & Rawson, 2009) and the desirable difficulties framework (Bjork, 1994) predict that difficult retrieval practice may benefit later performance greater than

easier retrieval practice. According to the retrieval effort hypothesis and the desirable difficulties framework, successful retrieval that is difficult will be more potent for memory and long-term performance compared to successful retrieval that is easy. Related to the present study, quizzing with higher order questions may be more challenging than quizzing with concept questions, and the retrieval effort engaged during higher order questions may overlap with the retrieval effort required at final test – potentially producing the largest retrieval practice effect.

Kang, McDermott, and Roediger (2007, Experiment 2) found results consistent with the transfer appropriate processing framework (also consistent with the retrieval effort hypothesis and desirable difficulties framework): retrieval practice with short answer quizzes (presumed to be more difficult in terms of retrieval effort) benefitted final performance to a greater extent than retrieval practice with multiple-choice quizzes, regardless of whether the final test was short answer or multiple-choice (when the initial quizzes were followed by feedback). Kang et al. explained that short answer tests might engage deeper recollective processing and greater retrieval effort compared to multiple-choice tests, thereby enhancing both short answer and multiple-choice delayed performance.

Regarding the current study, initial higher order retrieval practice may enhance both concept and higher order delayed performance, if higher order retrieval practice is challenging and engages retrieval effort required for both conceptual and higher order processing. In a study conducted at the same middle school as Experiment 3 from the present study was conducted, McDaniel

et al. (2011) demonstrated that retrieval practice via quizzing with application questions enhanced performance on both definition and application final tests, compared to no quizzing. These findings support the notion that retrieval practice with higher order (application) questions can benefit both lower order (definition) and higher order (application) delayed performance in an applied setting.

As stated earlier, it is commonly believed that in order to engage in higher order processing, students must first have an understanding of factual and conceptual knowledge (i.e., the "thinking with the basics" viewpoint). Accordingly, if retrieval practice with concept tests engages processes required for both concept and higher order skills, then both concept and higher order delayed performance may be enhanced. Supporting this prediction, McDaniel et al. (2011) also found that retrieval practice with definition questions enhanced performance on both definition and application final tests (though to a lesser extent in McDaniel et al.'s third experiment), compared to no quizzing.

In sum, the transfer appropriate processing framework suggests both "congruent" effects of retrieval practice where initial quiz and final test formats match (e.g., conceptual retrieval practice may enhance delayed concept performance and higher order retrieval practice may enhance delayed higher order performance) as well as "incongruent" effects of retrieval practice (e.g., conceptual retrieval practice may enhanced delayed higher order performance, and higher order retrieval practice may enhance delayed concept performance).

Note that I continue to use the term “congruent” to refer to the concept quiz-concept test and higher order quiz-higher order test conditions where initial quiz and final test formats match, and also the term “incongruent” to refer to the concept quiz-higher order test and higher order quiz-concept test conditions, where there is a mismatch between initial quiz and final test formats. For the “incongruent” conditions in particular, although there is a disparity in test formats, bear in mind that processing across the quizzes and tests may, indeed, overlap (hence a benefit according to the transfer appropriate processing framework). Still, it is important to note that students do not always transfer their knowledge from one situation to another (e.g., Bransford, Sherwood, Vye, & Rieser, 1986; Gick & Holyoak, 1980). In particular, students’ factual and conceptual knowledge may remain “inert” and students may fail to transfer this knowledge when answering higher order questions (or vice versa). The present study is designed to provide some insight regarding retrieval practice and its potential benefit for congruent versus incongruent learning, particularly the relationship between basic fact or concept learning and higher order learning.

Cognitive Load Theory

In contrast to beliefs held by educators (Kohn, 1999; Kuhn, 2005), some cognitive psychologists argue that in order to foster higher order skills, teachers must focus on and reinforce basic fact and concept learning (Willingham, 2009). Willingham argued that retrieval practice facilitates automatic recall of facts, requiring less effort and capacity from working memory, thereby enabling the student to use the additional working memory capacity for more complex skills.

Consistent with Willingham's (2009) argument, cognitive load theory posits that learning can be impaired when cognitive demands (imposed by a task) exceed the limited capacity of our working memory systems (Sweller, 1994; Plaas, Moreno, & Brünken, 2010); when demands increase past a certain level, learning generally decreases. On the other hand, if cognitive demands are reduced or diminished, learning may increase.

More specifically, Sweller (2010) explained that novel information is originally processed by working memory, and because working memory is limited in terms of capacity and duration, materials to be learned impose a working memory or cognitive load. According to cognitive load theory, there are three types of cognitive load: intrinsic, extrinsic, and germane. Intrinsic load is imposed by difficult material that requires a great deal of simultaneous processing (e.g., solving an algebraic equation); this load varies across individuals and types of material. Extrinsic load is imposed by difficult instructional techniques that add additional processing requirements (e.g., discovery learning). Germane load occurs when resources are devoted to schema acquisition (e.g., via retrieval practice) and this type of load enhances learning (Paas, Renkl, & Sweller, 2003).

A balance between intrinsic, extrinsic, and germane load must be met in order to maximize learning. It is not the case, for instance, that the more load imposed, the worse learning outcomes. Instead, if intrinsic load and extrinsic load are minimized while germane load is increased, then learning should be optimized. Specifically, working memory resources used to address extraneous

cognitive load should be minimized, thereby freeing up resources that can be reallocated to intrinsic and germane load in order to maximize learning. By increasing germane load (via schema acquisition), intrinsic load is reduced, freeing working memory capacity for additional schema acquisition (Paas, Renkl, & Sweller, 2003). This cycle of schema acquisition continues, facilitating the learning of increasingly complex knowledge and skills (see also Willingham, 2009, Chapter 5).

Thus, according to cognitive load theory for the current study, retrieval practice with concept questions should strengthen a student's knowledge base and increase germane load (schema acquisition and automation), thereby reducing intrinsic load and facilitating performance on both the delayed concept test and the delayed higher order test. Retrieval practice with higher order questions may also increase germane load; however, it is less clear from cognitive load theory how schema acquisition can be facilitated via difficult and complex questions. Instead, higher order questions may increase intrinsic load, rather than decrease it, impairing delayed performance.

Various techniques can be used to measure cognitive load, but they are still under development; for instance, techniques developed thus far generally measure cognitive load without specifying among intrinsic, extraneous, and germane load (see Brünken, Seufert, and Paas, 2010, for a review of measurement techniques). Often, cognitive load is measured using a dual task paradigm, whereby subjects engage in two tasks simultaneously and dual task performance is compared to single task performance. The reduction in

performance during a dual task is a measure of the cognitive load imposed by the additional task. In addition to cognitive load measures via dual task paradigms, measurements of eye movements and heart rates can also be used.

Another method that has been developed to measure cognitive load is the use of subjects' mental effort ratings (Paas, 1992; Tuovinen & Paas, 2004). While subjective, this type of measurement has been used in research on multimedia learning (Mayer & Moreno, 2003, 2010, for reviews), problem solving (Paas & van Merriënboer, 1994; Renkl & Atkinson, 2003, 2010, for reviews), and computer simulation tasks (Kester, Paas, & van Merriënboer, 2010, for a review) to develop various instructional techniques that reduce cognitive load and improve learning. In addition, Paas, van Merriënboer, and Adam (1994) demonstrated that the mental effort rating technique is both reliable (Cronbach's alpha between .82-.90) and sensitive to task differences when completing a variety of mathematical transfer problems (see Paas, 1992, for examples). At the same time, it remains unclear whether mental effort ratings can be used to measure a specific type of load (intrinsic, extrinsic, or germane) and also whether these ratings correlate with other cognitive load measures (DeLeeuw & Mayer, 2008).

This is the first investigation into the role retrieval practice may play in reducing cognitive load, and given the continuing development of cognitive load measures, the use of mental effort ratings in the current study is exploratory. Mental effort ratings were used in the current study in order to maintain some aspects of an educational setting to a greater extent than a dual task paradigm

would allow. Although this measurement technique is subjective and exploratory, it may provide some insight regarding the potential for concept learning to reduce cognitive load, a claim commonly asserted by cognitive psychologists (e.g., Willingham, 2009) that lacks empirical evidence.

For the current study, in Experiment 1, mental effort ratings were compared for four conditions: 1) retrieval practice with concept questions, delayed performance on concept questions (concept-concept), 2) retrieval practice with higher order questions, delayed performance on higher order questions (higher-higher), 3) retrieval practice with concept questions, delayed performance on higher order questions (concept-higher), and 4) retrieval practice with higher order questions, delayed performance on concept questions (higher-concept). Regarding the first two congruent conditions, mental effort ratings were expected to decrease across the two-day delay due to an increase in performance via retrieval practice, accompanied by a decrease in load expended.

Regarding the incongruent conditions (3 and 4, above), cognitive load theory predicts that mental effort ratings and delayed test performance for the concept-higher condition should decrease and delayed test performance should increase because retrieval practice with concept questions should decrease the amount of cognitive load required, subsequently enhancing delayed performance. For the higher-concept condition, on the other hand, initial mental effort required may be high and test phase mental effort may also be high because subjects may lack conceptual knowledge, subsequently impairing

delayed performance because cognitive demands may exceed subjects' capacity. Therefore, according to cognitive load theory, it was expected that the mental effort ratings would be greater for the higher-concept condition than for the concept-higher condition after a delay.

Again, the use of mental effort ratings in the present study was exploratory, although this measure was included to provide some insight into 1) the relationship between fact/concept learning and higher order learning, and 2) whether retrieval practice improves performance and accordingly reduces cognitive load over time.

Summary of Predictions

To summarize, the transfer appropriate processing framework posits that retrieval practice should enhance delayed performance when the initial question format matches the final question format or when initial processing matches the type of final processing engaged, compared to restudying. Cognitive load theory specifically predicts that retrieval practice with concept questions should benefit delayed concept and higher order performance via increased germane load and decreased intrinsic load. Finally, retrieval practice with higher order questions may benefit delayed concept performance if higher order questions present a desirable difficulty and engage conceptual (transfer appropriate) processing, but not if the cognitive load required is too great.

Introduction to Experiments

The current study includes three experiments that investigated how retrieval practice, typically used to enhance fact learning, can be used as a

strategy to improve higher order skills in both laboratory and applied settings. For the present purpose, higher order skills are considered to comprise the *understand*, *apply*, *analyze*, *evaluate*, and *create* categories of a revised Bloom's taxonomy (Anderson et al., 2001).

Across three experiments, subjects engaged in retrieval practice with only concept questions, only higher order questions, or a mix of question types. Performance was measured after a two-day delay on both concept and higher order questions in order to determine the type of retrieval practice that produces the largest gain in delayed performance.

The principal aim of Experiment 1 was to answer two questions: First, since we know that retrieval practice is a beneficial strategy for improving fact learning, does improved fact learning subsequently benefit higher order skills? Second, can retrieval practice directly promote higher order skills via critical thinking and analysis test questions? For Experiment 2, the principal aim was to evaluate how much time should be spent on concept vs. higher order learning – is a "mix" of the two question types used during retrieval practice, or the use of one type of question, more beneficial for enhancing higher order skills? Finally, the principal aim for Experiment 3 was to extend these findings to a middle school classroom under realistic conditions – textbook materials, 6th grade students, and in-class quizzes and exams. Across the three experiments, if retrieval practice benefits *both* concept learning and higher order skills, educators may be more willing to adopt this strategy in their classrooms.

Experiment 1

Experiment 1 was designed to accomplish two goals. First, can retrieval practice with concept questions improve delayed higher order performance? In other words, is it the case that higher order thinking skills are influenced, or even enhanced, when concept learning is improved (relative to no testing or restudying)? And second, can retrieval practice with higher order questions directly improve delayed higher order performance? Prior research examining retrieval practice has included some higher order questions, but often in the lower categories of Anderson et al.'s (2001) revised taxonomy, such as the *understand* and *apply* categories. Instead, if retrieval practice with higher order questions improves delayed higher order performance, retrieval practice may serve as a flexible and potent strategy for teachers and students in improving various types of learning.

In Experiment 1, subjects participated in four initial learning conditions: they studied a passage once, they studied a passage twice, they studied a passage once followed by a concept quiz, and they studied a passage once followed by a higher order quiz. After two days, subjects completed both concept and higher order tests for each condition, while also making mental effort ratings during both initial and final sessions to provide a measure of cognitive load.

First, it was predicted that retrieval practice (regardless of question type) would improve both delayed concept and higher order test performance more than studying once or twice (without quizzes), consistent with past research demonstrating the mnemonic benefits of retrieval practice compared to

restudying (e.g., Carrier & Pashler, 1992; Roediger & Karpicke, 2006a). In addition, initial and delayed performance was predicted to be greater for concept questions than for higher order questions, due to item or material difficulty. Foremost, it was predicted that retrieval practice with concept questions would benefit delayed higher order test performance and result in reduced mental effort ratings compared to restudying, consistent with cognitive load theory. However, it was less clear whether retrieval practice with higher order questions would benefit delayed concept test performance (compared to restudying), particularly if the cognitive demands during initial learning were too great.

Method

Participants. Forty-eight college students (*M* age = 20.58 years, 29 females) were recruited from the Department of Psychology human subject pool. Subjects received either credit towards completion of a research participation requirement or cash payment ($25). Analyses were conducted only after data from 48 subjects were collected, a sample size determined at the outset of the study using a power analysis with an assumed effect size of $d = 0.5$.

Design. A 4 x 2 within-subject design was used, such that four initial learning conditions [study once, study twice, concept quiz (1X), higher order quiz (1X)] were crossed with two delayed test types (concept, higher order). Eight passages, two per initial learning condition, were presented in the same order for all subjects, but the order in which the conditions occurred was blocked by learning condition and counterbalanced using a Latin Square (see Appendix A for the counterbalancing order). Learning conditions appeared once in every ordinal

position and were crossed with two types of final test questions, creating eight counterbalancing orders. Six subjects were randomly assigned to each of the eight orders. After a two-day delay (i.e., 48 hours later), subjects completed one test type (concept or higher order) per passage, with tests presented in the same order as passages were encountered during Session 1, and with questions blocked by passage.

Materials. Eight passages were adapted from eight different books included in the "Taking Sides" McGraw-Hill Contemporary Learning Series (http://www.mhcls.com). Each passage was approximately 1,000 words in length (*M* = 1,006 words, range = 990 to 1016 words), with half of each passage presenting one side of a debate, and the remaining half of each passage presenting the opposite side of the debate (see Appendix B for all passages). For example, the "Does welfare do more harm than good?" passage was adapted from *Taking sides: Clashing views on controversial social issues* (Finsterbusch & McKenna, 1984). Approximately 500 words were drawn from the book to describe the "yes" side of the argument, and approximately 500 words were used to describe the "no" side of the argument.

For Session 1, eight four-alternative multiple-choice concept questions and eight four-alternative multiple-choice higher order questions were developed for each passage (see Appendix C for Session 1 questions). For each question type, approximately four questions pertained to the "yes" side of the passage debate, and approximately four questions pertained to the "no" side of the debate. For Session 2, all question stems were rephrased and multiple-choice

alternatives were randomly re-ordered (see Appendix D for Session 2 questions). Across Session 1 and Session 2, regardless of counterbalancing order, the correct multiple-choice alternative appeared in every position (1, 2, 3, or 4) an equal number of times.

For the concept questions, key ideas stated in the passages were tested in order to measure subjects' basic understanding of the content. For example, a concept question from the "Does welfare do more harm than good?" passage included:

Which is the primary reason the "yes" author is against welfare programs?

1) Welfare programs don't benefit recipients or taxpayers
2) Welfare programs create dependence for recipients
3) Welfare programs are too expensive for taxpayers
4) Welfare programs are not the government's responsibility

The correct answer for the concept question is alternative #1, and this answer was stated directly in the passage. All concept questions in the present study were designed to encompass key concepts or ideas from passages, rather than details such as names, dates, vocabulary words, definitions, etc.

The higher order questions were developed in accordance with the *apply*, *analyze*, *evaluate*, and *create* categories of Anderson et al.'s (2001) revised Bloom's taxonomy (see Figure 1). Questions classified within these four categories were used to engage a variety of higher order skills, although the categories were not evaluated as an independent variable due to the limited number of items per category and passage. Rather, all *apply*, *analyze*, *evaluate*,

and *create* items used in the current study were considered to engage higher order skills.

For *apply* questions, subjects were asked a question about a new situation or problem that was related to a situation or problem stated in the passage (these items might also be called transfer items). For example, an *apply* question from the same passage about welfare included:

> What type of society would the "yes" author expect if there were no welfare programs in the future?
>
> 1) A society in which all individuals are self-reliant and independent
> 2) A society in which there would be no role for the government
> 3) A society in which no one would be required to pay taxes
> 4) A society in which all individuals are treated equally

The correct answer for the *apply* question is alternative #1, and this answer could be inferred from information stated in the passage.

For *analyze* questions, subjects were asked to differentiate the authors' arguments; they were presented with a statement and asked which author (the "yes" author, the "no" author, both authors, or neither author) would agree or disagree with the statement. For example, an *analyze* question included:

> Which author would agree with the following statement? "It is honorable for the government to help society."
>
> 1) The "yes" author
> 2) The "no" author

3) Both authors
4) Neither author

The correct answer for the *analyze* question is alternative #3.

For *evaluate* questions, subjects were asked to check or critique an author's argument by selecting a statement (which was not presented in the passage) that most accurately summarized the author's argument. For example, an *evaluate* question included:

> Which statement is an accurate evaluation or summary of the "yes" author's views?
>
> 1) Welfare programs can never work, because they are always too expensive
> 2) Welfare programs are harmful, because they make bad situations even worse
> 3) Welfare programs waste taxpayer money on people who don't really need help
> 4) Welfare programs could work, but they rarely meet the needs of the people

The correct answer for the *evaluate* question is alternative #4.

And finally, for *create* questions, subjects were asked to plan or predict an outcome for a novel situation that was not stated in the passage; thus, the author's reaction must be generated based on information presented in the passage. For example, a *create* question included:

How do you predict the “yes” author would react if he or she became unemployed and needed welfare assistance?

1) The “yes” author might accept government assistance, but would seek help from local organizations first
2) The “yes” author would not accept government assistance, but would try to find a new job
3) The “yes” author might accept government assistance, but would try to find a new job first
4) The “yes” author would not accept government assistance, but would seek help from local organizations

The correct answer for the *create* question is alternative #2. Again, all test questions used during Sessions 1 and 2 can be found in Appendices C and D.

Procedure. Subjects were tested in small groups (up to five people) using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2007). At the beginning of Session 1, subjects were instructed that they would be reading passages and taking multiple-choice tests. Subjects were presented with a sample passage about the Nicaraguan Contras (see Appendix B for the sample passage) for 20 seconds and they were instructed to try the computer program’s scrolling feature (viewing the entire passage using the up and down keys on the keyboard) without worrying about reading the passage. Next, subjects were presented with a sample test of two 4-alternative multiple-choice questions (self-paced) that asked subjects whether they turned off their cell phone and whether they could return in two days; in other words, subjects did not receive sample test

questions related to the sample passage. After responding to each sample multiple-choice question, subjects were asked to make a mental effort rating on a nine-point scale (described below), which was followed by 10 seconds of feedback, in order to acclimate subjects to the actual procedure.

After the instruction phase during Session 1, subjects completed two blocks: first, subjects read all eight passages (in the same order for all subjects), and second, subjects completed four conditions (two per passage), blocked by learning condition (order counterbalanced across subjects, see Appendix A). In other words during the second block, subjects did not re-encounter two of the passages (in the study once condition), they read two of the passages for a second time (in the study twice condition), they completed two quizzes with concept questions on two of the passages, and they also completed two quizzes with higher order questions on the remaining two passages.

During six-minute study periods, each passage was presented in its entirety on the computer screen and subjects were able to scroll up and down in order to read the complete text at their own pace. Subjects were asked to study the passage during the time allotted, and after six minutes the computer moved on to the next passage (during the first reading block) or to the appropriate condition (during the second condition block following the restudy condition). Subjects' keyboard presses were recorded during study periods, to ensure that all subjects scrolled appropriately through the passages from the top of the passage to the bottom.

During self-paced test periods, multiple-choice questions (blocked by passage) were presented one at a time, in a different random order for each subject. Subjects were asked to type a number (1, 2, 3, or 4) corresponding to the multiple-choice alternative (forced choice). As soon as subjects responded to each question, the computer moved on (i.e., subjects were not allowed to change their answer) and subjects were asked to estimate the mental effort required ("How much mental effort did you invest in this question") on a 9-point scale (adapted from Paas, 1992) by typing a number corresponding to the rating. The rating scale was as follows: 1) extremely low mental effort, 2) very low mental effort, 3) low mental effort, 4) somewhat low mental effort, 5) neither low nor high mental effort, 6) somewhat high mental effort, 7) high mental effort, 8) very high mental effort, 9) extremely high mental effort.

After subjects rated their mental effort, the computer presented feedback for 10 seconds by displaying the word "CORRECT" or "INCORRECT" corresponding to subjects' response, while also displaying the original question and the correct answer (without incorrect multiple-choice lures). After 10 seconds, the computer moved on to the next question. In other words, multiple-choice question responses and mental effort ratings were self-paced, whereas feedback was experimenter-controlled and presented for 10 seconds per item. During testing, subjects completed a question, mental effort rating, and then viewed feedback, followed by the next question.

After each passage and test (regardless of condition), subjects received a 15 second break, during which the computer screen displayed, "Please clear

your mind and wait for the computer to move on." Then, the computer moved on to the next passage or condition, according to subjects' counterbalancing order.

After two days, subjects returned for Session 2 and completed multiple-choice concept tests for four of the passages and multiple-choice higher order tests for the other four passages. Testing procedures outlined above for Session 1 were followed during Session 2 (including mental effort ratings and 15-second breaks), except subjects did not receive feedback during Session 2.

In sum, subjects participated in only four within-subject learning conditions, crossed with two delayed test types. Dependent variables collected included only accuracy on test questions, response times for test questions, mental effort ratings for test questions, and response times for mental effort ratings (not reported here). The entire procedure lasted approximately two and a half hours across the two sessions. At the end of the experiment, subjects were debriefed and thanked for their time.

Results

All results in the current study were significant at an alpha level of .05. A Greenhouse-Geisser correction was applied to analyses of variance (ANOVAs) when the sphericity assumption was violated (Greenhouse & Geisser, 1959), and a Bonferroni correction for multiple comparisons was applied to *p* values from *t*-tests by multiplying the *p* value by the number of comparisons (Rice, 1989). Effect sizes reported include partial eta-squared (η_p^2) for ANOVAs (Pearson, 1911; Pierce, Block, & Aguinis, 2004) and Cohen's *d* for *t*-tests (Cohen, 1988). Error bars in figures represent 95% confidence intervals, specifically calculated

for within-subject designs according to methods described by Cousineau (2005) and Morey (2008).

Initial quiz performance, reaction time, and mental effort ratings. Initial performance on the concept quiz and on the higher order quiz is displayed in Table 1. As predicted, initial performance was greater on the concept quiz (59%) compared to performance on the higher order quiz (47%), likely due to item difficulty, confirmed by a one-way ANOVA on initial performance, $F(1, 47) = 12.62$, $p = .001$, $\eta_p^2 = .21$.

Average reaction time for questions answered correctly on initial quizzes is displayed in Table 2. Subjects' reaction time was faster for concept questions ($M = 16.5$ seconds) than for higher order questions ($M = 21.1$ sec), confirmed by a one-way ANOVA on reaction time, $F(1, 47) = 35.91$, $p < .001$, $\eta_p^2 = .43$. Consider that concept questions used in Experiment 1 (and Experiment 2) had an average word count of $M = 327$ (including the question stem and multiple-choice alternatives), substantially less than the word count for higher order questions, $M = 427$. Thus, it is probable that reaction time may have been faster on concept questions due to fewer words used for these items.

Average mental effort ratings for questions answered correctly on initial quizzes are displayed in Table 3. Subjects' mental effort ratings (on a 9-point scale, with higher ratings representing greater cognitive load) were similar for concept questions ($M = 4.36$) and higher order questions ($M = 4.50$), and this difference was not statistically significant, $F < 1$. Thus, although subjects had greater performance and faster reaction times for concept questions, they

reported similar levels of cognitive load for concept questions and higher order quiz questions.

Final test performance. Final test performance on rephrased questions for the four initial learning conditions is displayed in Table 1 and Figure 2. When collapsed over final test type, delayed performance was lower for the study once (49%) and study twice (51%) initial learning conditions, and greater for the concept quiz (1X, 62%) and higher order quiz (1X, 62%) learning conditions. Consistent with initial quiz performance, overall performance on the final concept test (60%) was greater than for the final higher order test (53%), likely due to item difficulty. A 4 [initial learning condition: study once, study twice, concept quiz (1X), higher order quiz (1X)] x 2 (delayed test type: concept, higher order) repeated measures ANOVA on delayed performance indicated a main effect of learning condition, $F(3, 141) = 13.47$, $p < .001$, $\eta_p^2 = .22$, a main effect of delayed test type, $F(1, 47) = 12.47$, $p = .001$, $\eta_p^2 = .21$, and a significant interaction between learning condition and test type, $F(3, 141) = 27.39$, $p < .001$, $\eta_p^2 = .37$.

Regarding delayed performance on the concept test, post-hoc *t*-tests confirmed a significant effect of retrieval practice, such that final concept test performance for the concept quiz condition (1X, 78%) was significantly greater than final concept test performance for the study once (54%) and study twice (54%) conditions, $t(47) = 5.96$, $p < .001$, $d = 1.23$ and $t(47) = 6.63$, $p < .001$, $d = 1.24$, respectively. However, final concept test performance was similar for the higher order quiz condition (1X, 53%) compared to the study once and study twice conditions, ts < 1, indicating that retrieval practice with higher order

questions did not benefit final concept test performance (compared to no quizzing). In other words, an initial concept quiz improved final concept test performance (78%) to a much greater degree than an initial higher order quiz (53%), $t(47) = 6.93$, $p < .001$, $d = 1.29$. There was also no effect of restudying when comparing final concept test performance for the study once and study twice conditions, $t < 1$.

Regarding delayed performance on the higher order test, post-hoc *t*-tests also confirmed a significant effect of retrieval practice, such that final higher order test performance for the higher order quiz (1X) condition (72%) was significantly greater than final higher order test performance for the study once (44%) and study twice (49%) conditions, $t(47) = 8.17$, $p < .001$, $d = 1.39$ and $t(47) = 5.31$, $p < .001$, $d = 1.12$, respectively. However, final higher order test performance was similar for the concept quiz (1X) condition (46%) compared to the study once and study twice conditions, *t*s < 1, indicating that retrieval practice with concept questions did not benefit final higher order test performance (compared to no quizzing). In other words, an initial higher order quiz improved final higher order test performance (72%) to a much greater degree than an initial concept quiz (46%), $t(47) = 6.73$, $p < .001$, $d = 1.21$. Again, there was no effect of restudying on final higher order test performance when comparing the study once and study twice conditions, $t(47) = 1.40$, $p > .05$.

In sum, initial retrieval performance enhanced final test performance, but only when the initial quiz type (e.g., concept or higher order) matched the final test type (concept or higher order, respectively). For these congruent conditions,

performance was marginally greater for the concept quiz-concept test condition (78%) than for the higher order quiz-higher order test condition (72%), $t(47) = 1.94$, $p = .059$, $d = 0.32$, though this difference may be due to relative difficulty between concept versus higher order test questions. In contrast to my predictions, no benefit was found from initial concept retrieval practice on delayed higher order test performance, nor from initial higher order retrieval practice on delayed concept test performance.

Final reaction time. Subjects' average reaction time (RT) for questions answered correctly on the final test is displayed in Table 2. In general, subjects were slower on final higher order test questions ($M = 19.7$ sec per question) compared to final concept test questions ($M = 15.9$ sec), and consistent with final test performance patterns, subjects were slower in the study once ($M = 19.1$ sec) and study twice ($M = 19.2$ sec) conditions compared to the concept quiz (1X, $M = 17.2$ sec) and higher order quiz (1X, $M = 15.5$ sec) conditions. Three subjects were dropped from RT analyses because they did not answer any correct items in one or more conditions.

A 4 [initial learning condition: study once, study twice, concept quiz (1X), higher order quiz (1X)] x 2 (delayed test type: concept, higher order) repeated measures ANOVA on reaction time indicated a main effect of learning condition, $F(3, 132) = 7.10$, $p < .001$, $\eta_p^2 = .14$, a main effect of test type, $F(1, 44) = 40.26$, $p < .001$, $\eta_p^2 = .48$, and a significant interaction between learning condition and test type, $F(3, 132) = 6.24$, $p = .002$, $\eta_p^2 = .12$. For the interaction, subjects were fastest in the two congruent conditions, namely in the concept quiz-concept test

condition (*M* = 12.9 sec) compared to RT for other conditions on the final concept test, *t*s > 2.62, *p*s < .036, *d*s > 0.46, and also fastest in the higher order quiz-higher order test condition (*M* = 15.1 sec) compared to RT for the other conditions on the final higher order test, *t*s > 3.56, *p*s < .003, *d*s > 0.65. Consistent with the finding of lower final test performance for incongruent conditions, reaction time was slower for the incongruent conditions (concept quiz-higher order test and higher order quiz-concept test), not faster as originally predicted.

Final mental effort ratings. Average mental effort ratings for questions answered correctly on the final test are displayed in Table 3. When collapsed over test type, final mental effort ratings were greater for the study once (*M* = 5.05) and study twice (*M* = 4.74) conditions, compared to the concept quiz (1X, *M* = 4.35) and higher order quiz (1X, *M* = 4.28) conditions. This finding suggests that subjects had difficulty answering questions about passages that were studied once or twice (without quizzing), resulting in lower performance, slower reaction time, and greater cognitive load expended after a two-day delay. In addition, ratings during the final concept test (*M* = 4.48) were lower than during the final higher order test (*M* = 4.73), indicating that subjects expended more effort on higher order questions after a delay. Again, the same three subjects were dropped from mental effort rating analyses because they did not answer any correct items in one or more conditions.

A 4 [initial learning condition: study once, study twice, concept quiz (1X), higher order quiz (1X)] x 2 (delayed test type: concept, higher order) repeated

measures ANOVA on mental effort ratings indicated a main effect of learning condition, $F(3, 132) = 10.46$, $p < .001$, $\eta_p^2 = .19$, a main effect of test type, $F(1, 44) = 8.17$, $p = .006$, $\eta_p^2 = .16$, and a significant interaction between learning condition and test type, $F(3, 132) = 16.16$, $p < .001$, $\eta_p^2 = .27$.

Regarding mental effort ratings on the final concept test, the concept quiz condition resulted in the lowest mental effort rating (1X, $M = 3.72$) compared to the higher order quiz (1X, $M = 4.61$), study twice ($M = 4.68$), and study once ($M = 4.91$) conditions, respectively. Post-hoc *t*-tests confirmed that the concept quiz condition resulted in the lowest mental effort rating on the final concept test compared to the other three conditions, ts > 5.50, ps < .001, ds > 0.63, with no differences among the study once, study twice, and higher order quiz conditions, ts < 1.64. These results indicate that concept quizzing reduced cognitive load on a final concept test (consistent with increased test performance in the concept quiz-concept test condition), whereas restudying and higher order quizzing did not reduce cognitive load on the final concept test.

For mental effort ratings on the final higher order test, the higher order quiz condition resulted in the lowest mental effort rating (1X, $M = 3.96$) compared to the study twice ($M = 4.80$), concept quiz (1X, $M = 4.97$), and study once ($M = 5.18$) conditions, ts > 4.72, ps < .001, ds > 0.65, with no differences among the other three conditions, ts < 1.25, providing additional evidence that higher order quizzing benefitted higher order final test performance, and that subjects were sensitive to this benefit in their mental effort ratings. Similar to findings from final performance and final reaction time, a reduction in cognitive load for the

incongruent conditions was not found; on the contrary, cognitive load in these conditions was similar to the mental effort expended on the final test for study once and study twice conditions.

Discussion

In Experiment 1, retrieval practice with higher order questions greatly improved delayed higher order performance by 23-28% (compared to studying once or twice), and consistent with prior research, retrieval practice with concept questions also improved delayed concept performance by 24%. When the type of initial quizzing matched the type of final test, even when final test questions were rephrased, retrieval practice yielded comparable benefits on performance, reaction time, and mental effort ratings for both concept and higher order learning.

Contrary to cognitive load theory, retrieval practice with concept questions did not enhance delayed higher order performance. Similarly (and contrary to the desirable difficulty framework), retrieval practice with higher order questions did not enhanced delayed concept performance. There were also no benefits from restudying on delayed performance, reaction time, or mental effort ratings, even when the first and second study periods were spaced over time (i.e., not massed restudy, see studies by Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Callender & McDaniel, 2009; Carrier & Pashler, 1992; Karpicke & Roediger, 2007; Roediger & Karpicke, 2006b; Wheeler, Ewers, & Buonanno, 2003).

In sum, although there were no benefits of initial concept quizzing on delayed higher order performance, or benefits of initial higher order quizzing on delayed concept performance, robust retrieval practice effects were found for both concept and higher order learning, with no benefit of restudying on delayed performance.

Experiment 2

Considering the debate between how much time should be spent on fact or concept learning versus time spent on higher order learning, Experiment 2 was designed to examine whether a mix of both concept and higher order question types used during retrieval practice would be more beneficial for enhancing delayed higher order performance than using one type of question during initial quizzing. Also, considering the results from Experiment 1 where concept quizzing did not enhance delayed higher order performance and higher order quizzing did not enhance delayed concept performance, it may be the case that providing subjects with both types of questions during initial learning may benefit both types of learning, perhaps in an additive manner. For instance, when receiving both a concept quiz and a higher order quiz (i.e., in the mixed condition) in close succession during the first session, subjects may be more likely to extend their knowledge and improve subsequent learning compared to transferring knowledge from an initial concept quiz to a delayed higher order test across a two-day delay, a procedure that did not benefit delayed performance in Experiment 1.

In Experiment 2, subjects participated in four initial learning conditions, each after studying a passage once: they completed one higher order quiz; they completed two higher order quizzes; they completed two concept quizzes; and they completed a "mix" of two quizzes, one concept and one higher order. After two days, subjects completed both concept and higher order tests for each condition, while also making mental effort ratings during initial and final learning sessions to provide a measure of cognitive load. Passages studied once or twice (without quizzing) were not included in Experiment 2. Instead, comparisons of interest were between one higher order quiz vs. two higher order quizzes, and also the optimal combination of two quizzes for improving delayed concept and higher order learning – namely, two concept quizzes, two higher order quizzes, or a mix of quizzes.

Given the results from Experiment 1, it was expected that the higher order quiz (1X) and higher order quiz (2X) conditions would benefit delayed higher order performance more than delayed concept performance, and that the concept quiz (2X) condition would benefit delayed concept performance more than delayed higher order performance. Regarding one quiz versus two quizzes, it was predicted that two higher order quizzes would provide an additional benefit to delayed higher order performance compared to one higher order quiz. On the other hand, this additional benefit may be due to re-exposure to the same item twice; i.e., question stems were only rephrased between the initial and final sessions, not between the first and second initial quizzes.

Finally, regarding the mixed quiz condition (2X, with one concept and one higher order quiz), it may be the case that both quiz types provide subjects with "the best of both worlds," namely that the mixed quiz condition could enhance both types of delayed performance compared to the other two-quiz conditions (concept only or higher order only). In line with the transfer appropriate processing framework, engaging in both types of processing during initial learning may have the greatest overlap in processing to the two final test types, enhancing delayed performance. Cognitive load theory also suggests that if subjects engage in both types of processing during initial learning, it may reduce cognitive load (measured by subjects' mental effort ratings), thereby increasing performance on both of the final test types. At the same time, one quiz of each format (in the mixed quiz condition) may not prove as potent as two quizzes of the same format [in the concept (2X) and higher order (2X) quiz conditions]; therefore, it was unclear whether the mixed quiz (2X) condition would provide a smaller or larger benefit to delayed performance compared to the concept (2X) and higher order (2X) quiz conditions.

Method

Participants. Forty-eight college students (*M* age = 20.04 years, 31 females) were recruited from the Department of Psychology human subject pool. Subjects received either credit towards completion of a research participation requirement or cash payment ($25). Subjects who participated in Experiment 2 had not participated in Experiment 1. Analyses were conducted only after data

from 48 subjects were collected, a sample size determined at the outset of the study using a power analysis with an assumed effect size of $d = 0.5$.

Design. A 4 x 2 within-subject design was used, such that four initial learning conditions [higher order quiz (1X), higher order quizzes (2X), concept quizzes (2X), mixed quizzes (2X)] were crossed with two delayed test types (concept, higher order). Eight passages, two per initial learning condition, were presented in the same order for all subjects, but the order in which the conditions occurred was blocked by learning condition (as well as test type in Session 1) and counterbalanced using a Latin Square (see Appendix A for the counterbalancing orders). Learning conditions appeared once in every ordinal position and were crossed with two types of final tests, creating eight counterbalancing orders. Six subjects were randomly assigned to each of the eight orders.

For the mixed quiz (2X) condition, subjects completed a concept quiz followed by a higher order quiz, or they completed a higher order quiz followed by a concept quiz. Order of quizzes in the mixed quiz condition was counterbalanced equally across subjects (see Appendix A for the counterbalancing order).

After a two-day delay (i.e., 48 hours later), subjects completed one test type (concept or higher order) per passage, with tests presented in the same order as passages were encountered during Session 1, and with questions blocked by passage.

Materials. The same materials from Experiment 1 were used in Experiment 2.

Procedure. The same procedures used in Experiment 1 were used in Experiment 2, except that after the instruction phase during Session 1, subjects completed three blocks: first, subjects read all eight passages (in the same order for all subjects); second, subjects completed the first quiz block with eight quizzes (one quiz per passage, presented in the same order as passages during the reading block); and third, subjects completed a second quiz block with six quizzes [one quiz per passage, except for passages in the higher order quiz (1X) condition, again presented in the same order]. After two days, subjects returned for Session 2 and completed multiple-choice concept tests for four of the passages and multiple-choice higher order tests for the other four passages.

In sum, subjects participated in only four within-subject learning conditions, crossed with two delayed test types. Dependent variables collected included only accuracy on test questions, response times for test questions, mental effort ratings for test questions, and response times for mental effort ratings (not reported). The entire procedure lasted approximately two and a half hours across the two sessions. At the end of the experiment, subjects were debriefed and thanked for their time.

Results

Initial quiz performance. Initial performance during the first and second quiz blocks is displayed in Table 4. Regarding counterbalancing order for the mixed quiz (2X) condition, subjects who took a higher order quiz first and a

concept quiz second had similar initial performance (53%, collapsed over the two test types) compared to subjects who took a concept quiz first and a higher order quiz second (52%). This between-subject difference on initial performance was not significant, $F < 1$; thus, initial mean performance for the mixed quiz (2X) condition has been collapsed across the counterbalancing order (but see Table 4 for the complete set of means). In other words, means reported for the mixed quiz (2X) condition in the first quiz block include data from subjects whose first test was concept or higher order, and means reported for the second quiz block also include data from subjects whose second test was concept or higher order.

For the first quiz block, initial performance was greatest for the concept quiz (2X) condition (57%), followed by initial performance for the mixed quiz (2X, 52%, collapsed over test type), higher order quiz (2X, 49%), and higher order quiz (1X, 47%) conditions, respectively. For the second quiz block, initial performance was again greatest for the concept quiz (2X) condition (91%), followed by the higher order quiz (2X, 83%) and mixed quiz (2X, 53%, collapsed over test type) conditions.

A 3 [learning condition: higher order quiz (2X), concept quiz (2X), mixed quiz (2X)] x 2 (quiz block: first, second) repeated measures ANOVA on initial performance revealed a significant main effect of learning condition, $F(2, 94) = 64.27$, $p < .001$, $\eta_p^2 = .58$, a significant main effect of quiz block, $F(1, 47) = 356.69$, $p < .001$, $\eta_p^2 = .88$, and a significant interaction between learning condition and quiz block, $F(2, 94) = 42.77$, $p < .001$, $\eta_p^2 = .48$. As can be seen from Table 4, the higher order quiz (2X) and concept quiz (2X) conditions

resulted in a similar increase in performance from the first quiz block to the second quiz block (34% for each condition).

Performance in the mixed quiz (2X) condition, on the other hand, remained relatively constant across quiz blocks (see Table 4); keep in mind that performance for each quiz block includes subjects' performance on both types of quizzes (concept and higher order). Still, this finding suggests a replication of Experiment 1, namely that retrieval practice on one quiz format did not benefit performance on a second quiz of a different format, even in close succession during the first session – performance on the second quiz in the mixed condition was similar to performance on the first quiz of the same type in the concept quiz (2X) and higher order quiz (2X) conditions.

Post-hoc comparisons confirmed that the concept quiz (2X) condition resulted in greater performance than the higher order quiz (1X) and higher order quiz (2X) conditions on the first quiz block, $t(47) = 4.00$, $p < .001$, $d = 0.71$ and $t(47) = 2.66$, $p = .011$, $d = 0.56$, respectively, but concept quiz (2X) performance was not significantly greater than mixed quiz (2X) performance on the first quiz block, $t(47) = 1.91$, $p > .05$ (likely because the mixed quiz condition includes subjects whose first quiz was also a concept quiz). On the second quiz block, the concept quiz (2X) condition resulted in greater performance than the higher order quiz (2X) and mixed quiz (2X) conditions, $t(47) = 4.29$, $p < .001$, $d = 0.77$ and $t(47) = 15.66$, $p < .001$, $d = 3.13$, respectively. In general, the concept quiz (2X) performance resulted in substantially greater performance during both the

first and second quiz blocks compared to the other conditions, probably due to a difference in item difficulty between concept and higher order questions.

Initial reaction time. Average reaction time for questions answered correctly on initial quizzes is displayed in Table 5. Similar to initial performance, there was no effect of counterbalancing order on reaction times in the mixed quiz (2X) condition, $F(1, 46) = 1.22$, $p > .05$, so order of test types was collapsed across subjects (but see Table 5 for all means).

For the first quiz block, subjects' reaction time was fastest for the concept quiz (2X) condition (M = 16.5 seconds), followed by reaction time for the mixed (2X, M = 18.3 sec), higher order quiz (1X, M = 19.8 sec), and higher order quiz (2X, M = 20.4 sec) conditions, respectively. For the second quiz block, subjects' reaction time was again fastest for the concept quiz (2X) condition (M = 8.2 sec), followed by reaction time for the higher order quiz (2X, M = 9.4 sec) and mixed quiz (2X, M = 16.3 sec) conditions.

A 3 [learning condition: higher order quiz (2X), concept quiz (2X), mixed quiz (2X)] x 2 (quiz block: first, second) repeated measures ANOVA on initial reaction time revealed a significant main effect of learning condition, $F(2, 94) = 53.95$, $p < .001$, $\eta_p^2 = .53$, a significant main effect of quiz block, $F(1, 47) = 94.94$, $p < .001$, $\eta_p^2 = .67$, and a significant interaction between learning condition and quiz block, $F(2, 94) = 30.69$, $p < .001$, $\eta_p^2 = .40$. As can be seen from Table 5, reaction time from the first quiz block to the second quiz blocked dropped by about 8-10 seconds for both the higher order quiz (2X) and the concept quiz (2X) conditions, whereas reaction time dropped by only two seconds for the mixed

quiz (2X) condition, since each quiz block includes subjects' performance on both types of quizzes (concept and higher order).

Post-hoc comparisons confirmed that reaction time was faster for the concept quiz (2X) condition compared to the higher order quiz (1X) and higher order quiz (2X) conditions on the first quiz block, $t(47) = 4.42$, $p < .001$, $d = 0.51$ and $t(47) = 4.88$, $p < .001$, $d = 0.56$, respectively. In addition, reaction time was faster for the concept quiz (2X) condition compared to the higher order quiz (2X) and mixed quiz (2X) conditions on the second quiz block, $t(47) = 3.34$, $p = .006$, $d = 0.44$ and $t(47) = 11.32$, $p < .001$, $d = 1.90$, respectively. Finally, reaction time for the higher order quiz (2X) condition during the second quiz block was faster than for the mixed quiz (2X) condition, $t(47) = 10.11$, $p < .001$, $d = 1.63$. Overall, initial reaction times were consistent with initial test performance in that reaction times were fastest and test performance was greatest for the concept quiz (2X) condition, regardless of initial quiz block. Similar to Experiment 1, the faster reaction times for the concept quiz (2X) condition could be a result of item difficulty (i.e., concept quiz items were easier than higher order quiz items) or due to fewer words (100 words on average) used for the concept quiz items.

Initial mental effort ratings. Average mental effort ratings for questions answered correctly on initial quizzes are displayed in Table 6. Similar to initial performance and reaction time, there was no effect of counterbalancing order on reaction times in the mixed quiz (2X) condition, $F(1, 46) = 3.23$, $p > .05$, so order of test types was collapsed across subjects (but see Table 6 for all means).

For initial mental effort ratings during the first initial quiz block, ratings were the lowest for the concept quiz (2X) condition (*M* = 4.25), followed by the mixed quiz (2X, *M* = 4.40, collapsed over counterbalancing order), higher order quiz (2X, *M* = 4.63) and higher order quiz (1X, *M* = 4.71) conditions, respectively. For the second quiz block, mental effort ratings were again lowest for the concept quiz (2X) condition (*M* = 2.41), followed by the higher order quiz (2X, *M* = 2.56) and mixed quiz (2X) conditions (*M* = 4.29), respectively.

A 3 [learning condition: higher order quiz (2X), concept quiz (2X), mixed quiz (2X)] x 2 (quiz block: first, second) repeated measures ANOVA on initial mental effort ratings revealed a significant main effect of learning condition, $F(2, 94) = 42.39$, $p < .001$, $\eta_p^2 = .47$, a significant main effect of quiz block, $F(1, 47) = 82.06$, $p < .001$, $\eta_p^2 = .64$, and a significant interaction between learning condition and quiz block, $F(2, 94) = 47.18$, $p < .001$, $\eta_p^2 = .50$. As can be seen from Table 6, mental effort ratings from the first quiz block to the second quiz block dropped by about two points for the concept quiz (2X) and higher order quiz (2X) conditions, whereas ratings remained relatively constant across quiz blocks for the mixed quiz (2X) condition – possibly due to the collapsing of counterbalance order in the mixed quiz condition, but ratings may have also remained constant in the mixed quiz (2X) condition because retrieval practice on one quiz did not reduce cognitive load on a second quiz of a different format.

Post-hoc comparisons for ratings during the first quiz block confirmed that ratings were lower for the concept quiz (2X) condition compared to the higher order quiz (1X) and higher order quiz (2X) conditions, $t(47) = 3.29$, $p = .012$, $d =$

0.42 and $t(47) = 3.15$, $p = .018$, $d = 0.33$, respectively. Other comparisons between ratings during the first quiz block were not significant, ts < 1.85. For the second quiz block, ratings were significantly greater in the mixed quiz (2X) condition compared to the higher order quiz (2X) condition, $t(47) = 9.02$, $p < .001$, $d = 1.45$, and also greater compared to the concept quiz (2X) condition, $t(47) = 9.63$, $p < .001$, $d = 1.61$, although the difference in ratings between the higher order quiz (2X) and concept quiz (2X) conditions for the second test block was not significant, $t(47) = 1.41$, $p > .05$.

In sum, subjects reported expending less effort for the concept quiz (2X) condition during the first quiz block (and slightly less effort in this condition during the second block), and they also reported expending more effort for the mixed quiz (2X) condition during the second test block. Thus, the overall pattern of initial mental effort ratings was somewhat consistent with greater initial test performance and faster reaction times for the concept quiz (2X) condition, in particular.

Final test performance. Final test performance for the four initial learning conditions is displayed in Table 4 and Figure 3. There was no effect of counterbalancing order on final test performance for the mixed quiz (2X) learning condition, $F < 1$, therefore means reported have been collapsed over initial order (but see Table 4 for all means).

As seen on the far right side of Table 4, delayed performance was greatest for the mixed quiz (2X) condition (75%), compared to the concept quiz (2X, 69%), higher order quiz (2X, 69%), and higher order quiz (1X, 65%)

conditions, respectively. Overall performance for the two test types, however, was similar: 69% correct on the final concept test and 70% correct on the final higher order test. A 4 [learning condition: higher order quiz (1X), higher order quizzes (2X), concept quizzes (2X), mixed quizzes (2X)] x 2 (delayed test type: concept, higher order) repeated measures ANOVA on delayed performance revealed a main effect of learning condition, $F(3, 141) = 4.85$, $p = .003$, $\eta_p^2 = .09$, and a significant interaction between learning condition and delayed test type, $F(3, 141) = 86.23$, $p < .001$, $\eta_p^2 = .65$.

Regarding delayed performance on the concept test, post-hoc *t*-tests confirmed that the concept quiz (2X) condition (90%) and the mixed quiz (2X) condition (78%) resulted in greater delayed performance compared to the higher order quiz (1X, 54%) and the higher order quiz (2X, 53%) conditions, *t*s > 6.10, *p*s < .001, *d*s > 1.21. The difference in delayed concept test performance between the concept quiz (2X) and mixed quiz (2X) conditions was also significant, $t(47) = 3.72$, $p = .006$, $d = 0.77$. In general, initial concept and initial mixed quizzing improved delayed concept performance, though two concept quizzes improved delayed performance to a greater extent than one concept quiz and one higher order quiz (i.e., the mixed quiz condition).

Regarding delayed performance on the higher order test, post-hoc *t*-tests confirmed that the higher order quiz (2X, 85%), higher order quiz (1X, 77%), and mixed quiz (2X, 71%) conditions resulted in greater delayed performance compared to the concept quiz (2X) condition (48%), *t*s > 5.80, *p*s < .001, *d*s > 1.24. The difference between the higher order quiz (2X) and the mixed quiz (2X)

conditions was also significant, $t(47) = 4.52$, $p < .001$, $d = 0.84$; however, neither of these two conditions differed significantly from the higher order (1X) condition, ps > .05.

Finally, the difference in delayed performance between the congruent conditions, namely delayed concept test performance for the concept quiz (2X) condition (90%) and delayed higher order test performance for the higher order quiz (2X) condition (85%), was not significant, $t(47) = 2.01$, $p > .05$, although performance was close to ceiling levels. In addition, the difference between the mixed quiz (2X) condition on the delayed concept test (78%) versus the delayed higher order test (71%) was marginally significant, $t(47) = 2.08$, $p = .088$, $d = 0.39$, though this was likely due to differences in item difficulty.

Consistent with Experiment 1, the congruent conditions (concept quizzes-concept test, higher order quizzes-higher order test) resulted in the greatest delayed performance. The mixed quiz (2X) condition produced reduced levels of performance on both delayed test types compared to the congruent conditions, suggesting that two quizzes of the same format are more potent for long-term learning than one quiz of each format. Interestingly, the concept quiz (2X) condition still did not benefit delayed higher order performance, even when compared to only one initial higher order quiz, providing further evidence that a boost in concept learning does not necessarily improve delayed higher order performance.

Final reaction time. Subjects' average reaction time (RT) for questions answered correctly on the final test is displayed in Table 5. Again, mixed

quizzing counterbalancing order did not interact with final reaction time, $F < 1$, thus order of initial quizzes was collapsed in the mixed quiz (2X) condition (but see Table 5 for all means).

In general, subjects were slower on final higher order test questions (M = 13.9 sec) compared to final concept test questions (M = 12.7 sec), and consistent with overall performance on the final test, subjects were fastest in the mixed quiz (2X) condition (M = 12.4 sec), followed by the higher order quiz (2X, M = 12.9 sec), concept quiz (2X, M = 13.9 sec), and higher order quiz (1X, M = 13.9 sec) conditions, respectively.

A 4 [learning condition: higher order quiz (1X), higher order quizzes (2X), concept quizzes (2X), mixed quizzes (2X)] x 2 (delayed test type: concept, higher order) repeated measures ANOVA on final reaction time revealed a main effect of learning condition, $F(3, 141) = 3.78$, $p = .018$, $\eta_p^2 = .08$, a main effect of test type, $F(1, 47) = 14.86$, $p < .001$, $\eta_p^2 = .24$, and a significant interaction between learning condition and test type, $F(3, 141) = 45.65$, $p < .001$, $\eta_p^2 = .49$. For the interaction, subjects were fastest in the two congruent conditions, namely in the concept quiz (2X)-concept test condition (M = 9.9 sec) compared to RT for other conditions on the final concept test, ts > 2.31, ps < .075, ds > 0.36, and also in the higher order quiz (2X)-higher order test condition (M = 11.2 sec) compared to RT for the other conditions on the final higher order test, ts > 2.76, ps < .024, ds > 0.41.

The mixed quiz (2X) condition resulted in faster reaction times compared to the higher order quiz (1X) and higher order quiz (2X) conditions on the final

concept test, *t*s > 4.19, *p*s < .001, *d*s > 0.74. For the final higher order test, the mixed quiz (2X) and the higher order quiz (1X) conditions resulted in similar reaction times, $t = 1.14$, $p > .05$; however, both resulted in faster reaction times than the concept quiz (2X) condition, *t*s > 5.05, *p*s < .001, *d*s > 0.78.

Finally, the congruent concept quiz (2X)-concept test condition resulted in faster reaction time compared to the congruent higher order quiz (2X)-higher order test condition, $t(47) = 2.67$, $p = .010$, $d = 0.37$, and the mixed quiz (2X)-concept test condition resulted in faster reaction time than the mixed quiz (2X)-higher order test condition, $t(47) = 3.79$, $p < .001$, $d = 0.60$, likely due to differences in item difficulty or word count. In general, the reaction time results support findings from final test performance, in that the congruent conditions produced the greatest level of performance and fastest reaction times, compared to the mixed quizzing or incongruent conditions.

Final mental effort ratings. Average mental effort ratings for questions answered correctly on the final test are displayed in Table 6. There was no effect of counterbalancing order in the mixed quiz (2X) condition on final mental effort ratings, $F < 1$, thus analyses have been collapsed over order (but see Table 6 for all means).

When collapsed over final test type, mental effort ratings were lowest for the mixed quiz (2X) condition ($M = 3.68$), followed by the concept quiz (2X, $M = 3.89$), higher order quiz (2X, $M = 3.97$), and higher order quiz (1X, $M = 4.14$) conditions, similar to overall patterns in performance and reaction time.

Consistent with similar levels of final test performance on the final concept and higher order tests, ratings during the final concept test (*M* = 3.98) were similar to ratings during the final higher order test (*M* = 3.86), indicating that subjects expended similar effort on the two test types after a delay. A 4 [learning condition: higher order quiz (1X), higher order quizzes (2X), concept quizzes (2X), mixed quizzes (2X)] x 2 (delayed test type: concept, higher order) repeated measures ANOVA on final mental effort ratings indicated a main effect of learning condition, $F(3, 141) = 5.03$, $p = .005$, $\eta_p^2 = .10$, and a significant interaction between learning condition and test type, $F(3, 141) = 53.88$, $p < .001$, $\eta_p^2 = .53$.

Regarding mental effort ratings on the final concept test, ratings were lower for the concept quiz (2X, *M* = 3.10) and mixed quiz (2X, *M* = 3.60) conditions, followed by the higher order quiz (1X, *M* = 4.57) and higher order quiz (2X, *M* = 4.66) conditions. Post-hoc *t*-tests confirmed that the concept quiz (2X) condition resulted in the lowest mental effort rating on the final concept test compared to the other three conditions, *t*s > 3.03, *p*s < .024, *d*s > 0.41. The mixed quiz (2X) condition also resulted in lower mental effort ratings on the final concept test than the higher order quiz (1X) and (2X) conditions, *t*s > 4.68, *p*s < .001, *d*s > 0.80, though the difference between these latter two conditions was not significant, $t < 1$. These results indicate that concept quizzing and mixed quizzing reduced cognitive load on a final concept test (consistent with an increase in final test performance), whereas higher order quizzing did not reduce cognitive load on the final concept test.

For mental effort ratings on the final higher order test, ratings were lower for the higher order quiz (2X, *M* = 3.13) and higher order quiz (1X, *M* = 3.72) conditions, followed by the mixed quiz (2X, *M* = 3.76) and concept quiz (2X, *M* = 4.85) conditions. The higher order quiz (2X) condition resulted in the lowest mental effort rating compared to the other conditions, *t*s > 4.07, *p*s < .001, *d*s > 0.50. The higher order quiz (1X) and the mixed quiz (2X) conditions resulted in lower mental effort ratings than the concept quiz (2X) condition, *t*s > 5.53, *p*s < .001, *d*s > 0.89, though the difference between the higher order quiz (1X) and mixed quiz (2X) conditions was not significant, *t* < 1. In sum, higher order and mixed quizzing reduced cognitive load on the final higher order test, although concept quizzing did not provide the same benefit on the final higher order test.

Discussion

In Experiment 2, retrieval practice with two higher order quizzes improved delayed higher order performance by an additional 8% compared to only one higher order quiz, and consistent with Experiment 1, when the type of initial quizzing (via two quizzes) matched the type of final test, retrieval practice yielded benefits on performance, reaction time, and mental effort ratings for both concept and higher order learning to a greater extent than one quiz of each format (in the mixed quiz condition). Contrary to cognitive load theory, but replicating Experiment 1, retrieval practice with concept questions did not enhance delayed higher order performance. Similarly, retrieval practice with higher order questions did not enhanced delayed concept performance. The findings from Experiment 2 provide further evidence that retrieval practice is the most powerful

when the type of questions used during initial learning are of the same complexity or taxonomic category (see Figure 1) as questions used during the final test.

Experiment 3

Experiment 3 was designed to address whether results from Experiments 1 and 2 would extend to an applied setting with different materials (Social Studies textbook chapters) and a different population of subjects (6^{th} grade middle school students). Importantly, in order to encourage teachers to use retrieval practice in their classrooms, Experiment 3 examined 1) whether retrieval practice could benefit delayed performance (compared to no quizzing), 2) whether higher order quizzing could benefit delayed higher order performance in an applied setting (a finding not yet empirically demonstrated), and 3) whether mixed quizzing would benefit delayed concept and higher order performance in an applied setting.

Previous research demonstrated that retrieval practice enhances delayed performance (compared to no quizzing) for middle school students (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel, Thomas et al., 2011). In these studies, however, quiz questions were comprised of science facts, definitions, and some application questions, not more complex higher order concepts from taxonomic categories such as *analyze* or *evaluate* (see Figure 1). Thus, an aim of Experiment 3 was to extend the findings from Experiments 1 and 2 (as well as findings from previous research) by examining the benefit of retrieval practice with higher order quiz questions in an applied setting.

In Experiment 3, subjects participated in two initial learning conditions (one per chapter of Social Studies material): they completed three higher order quizzes or they completed three "mixed" quizzes, which included both concept and higher order questions. Two days later, subjects completed a final test comprised of quizzed concept and higher order questions, which also included non-quizzed concept and higher order control questions (items counterbalanced across subjects).

In order to maximize power using the largest number of items per condition as possible while reducing classroom time required for the manipulation, a restudy condition and mental effort ratings were not included in this experiment (although prior research in the same Social Studies classroom demonstrated that retrieval practice enhanced delayed performance compared to a restudying exposure control; see Roediger, Agarwal, McDaniel, & McDermott, in press). In the present experiment, non-tested concept and higher order items were developed for each chapter in order to provide a control comparison for the two retrieval practice conditions, although exposure was not controlled.

Based on prior research (and the current Experiments 1 and 2), retrieval practice (regardless of quiz condition) was expected to enhance both delayed concept and higher order performance, compared to delayed performance on non-quizzed items. Based on findings from Experiment 2, the mixed quiz (3X) condition was predicted to enhance delayed concept performance compared to the higher order quiz (3X) condition, and the higher order quiz (3X) condition was predicted to enhance delayed higher order performance more than the mixed for

concept quiz (3X) conditions, due to overlap between initial and final processing. Furthermore, based on results from the prior experiments, it was expected that the incongruent (concept quizzes-higher order test or higher order quizzes-concept test) and mixed quiz (3X) conditions would improve delayed performance to a lesser (or perhaps similar) extent compared to the congruent conditions in which initial quiz and final test question format matched.

Method

Participants. All 142 6th grade students (*M* = 24 students in each of six classroom sections; 71 males, 71 females) from a Midwestern suburban middle school participated in Experiment 3. Subjects did not receive compensation for participating because the project was part of their typical classroom instruction. Assent from each student was obtained in accordance with guidelines from the Human Research Protection Office. Of the 142 students, 12 declined to include their data in the study (but these students still participated in all activities), and data from eight special education students were excluded from analyses because special education students received accommodations (e.g., additional study and quiz opportunities outside of class).

Design. A 3 x 2 within-subjects design was used, such that three learning conditions [higher order quizzes (3X), mixed quizzes (3X), non-quizzed] were crossed with two delayed test types (concept, higher order). Conditions were manipulated across two chapters of Social Studies material, with chapters presented in the same order for all subjects (as determined by the classroom teacher). Six classroom sections were split into two sets (e.g., set A and set B)

of three class sections each; i.e., periods 1, 3, 6 comprised Set A and periods 2, 5, and 7 comprised Set B. For the first chapter, Set A students completed three quizzes with higher order questions, whereas Set B students completed three quizzes with a mix of question types. For the second chapter, the learning conditions per set switched (see Appendix A for the counterbalancing order for Experiment 3). At the end of each chapter unit (approximately 7-8 school days in length; 48 hours after the third quiz), subjects completed a final test comprised of both question types (concept and higher order), with all questions presented in a different random order for each of the six classroom sections. In addition, final test question stems were identical to initial quiz question stems due to classroom constraints, although order of multiple-choice alternatives was randomized across all quizzes and final tests.

Materials. Two Social Studies textbook chapters (Russian Revolution and World War II from Banks et al., 1997), assigned by the classroom teacher, were used in this experiment (see Appendix E for the chapters). Each chapter was approximately 2,350 words in length (e.g., 2,335 words for Russian Revolution and 2,407 words for World War II). Students read each chapter from their Social Studies textbook, listened to approximately 7-8 lectures, and completed assignments and exams developed by the teacher. The teacher's assignments and graded exams focused on vocabulary terms (words bolded in the chapters in Appendix E), and they did not include the concept or higher order questions developed by the experimenter.

Twelve four-alternative multiple-choice concept questions and twelve four-alternative multiple-choice higher order questions were developed for each chapter and all questions and multiple-choice alternatives were approved by the classroom teacher (see Appendix F for questions). Across all quizzes and delayed tests, each classroom section received a different set of quizzed and non-quizzed items, and every item was quizzed (and not quizzed) at least once. In addition, for every quiz and test, each classroom section received a different random order of questions and the multiple-choice alternatives were randomly re-ordered. The correct multiple-choice alternative appeared in every position (A, B, C, or D) an equal number of times across quizzes and tests.

For the concept questions, key ideas stated in the chapters were tested in order to measure subjects' basic understanding of the content. For example, a concept question from the "Russian Revolution" textbook chapter included:

Why was Nicholas II forced to give up his role as tsar?

A) Because the Duma elected a new tsar

B) Because Stalin took over the government

C) Because his wife and children moved to Moscow

D) Because of angry protestors, soldiers, and railroad workers

The correct answer for the concept question is alternative D, and this answer was stated directly in the passage. Similar to Experiments 1 and 2, all concept questions in the present experiment were designed to encompass key concepts or ideas from the textbook chapters, rather than details such as names, dates, vocabulary words, definitions, etc.

The higher order questions were developed in accordance with the *apply*, *analyze*, and *evaluate* categories of Anderson et al.'s (2001) revised Bloom's taxonomy (see Figure 1), using the same operational definitions as those used in Experiment 1 (see pp. 26-29). Higher order questions from the taxonomic *create* category were not included in Experiment 3, due to concerns that 6th grade students may have had difficulty extending textbook concepts to completely novel situations. In addition, higher order questions from different taxonomic categories were not evaluated as an independent variable due to the limited number of items per category and chapter. For example, an *analyze* question from the same Russian Revolution chapter included:

> Which person would agree with the following statement? "Revolutions are hard to prevent."
>
> A) Alexander II
>
> B) Lenin
>
> C) Nicholas II
>
> D) Stalin

The correct answer for the *analyze* question is alternative C.

Quizzes comprised either eight higher order questions or a mix of concept and higher order questions (four of each type). For mixed quizzes, question type (concept or higher order) was blocked (and order was counterbalanced across classroom sections and quizzes), with questions presented in random order within question type block. Questions that were not tested on initial quizzes (a

non-quizzed control condition) were covered in the textbook chapter and also the teacher's lectures.

Final chapter tests comprised all multiple-choice concept and higher order questions (12 concept and 12 higher order questions per chapter). Final chapter test questions were the same as those from initial quizzes (i.e., questions were not rephrased) due to concern regarding floor effects, and all 24 items were presented in random order (not blocked by question type) on final chapter tests.

Procedure. Subjects completed initial quizzes individually via a clicker response system (Ward, 2007) in the classroom using a computer, projector, and projection screen at the front of the classroom. At the beginning of the experiment, subjects were instructed that they would be taking quizzes (via clickers, with which students were already familiar) and tests as part of a research study, and that their scores may or may not count for a grade. In actuality, subjects' individual scores were not factored into their individual grades; instead, students' overall classroom section scores counted towards a pizza party held at the end of the school year. The classroom section with the highest score on each quiz or test received five points towards the pizza party, and the classroom section with the second highest score on each quiz or test received four points towards the pizza party. (Other classroom assignments and exams similarly factored into students' pizza party points.)

During all clicker quizzes (pre-, post-, and review quizzes), multiple-choice questions (higher order only or mixed question types) were displayed on a projection screen at the front of the classroom one at a time, in a different

random order for each classroom section. The experimenter read the question stem and four multiple-choice alternatives aloud. After the experimenter was finished reading the question and alternatives, subjects were allowed to type a letter (A, B, C, or D) on their clicker remote corresponding to the multiple-choice alternative (forced choice). Once all subjects in the classroom responded (after approximately one minute), the experimenter closed the response option and the clicker software displayed the question stem, all four multiple-choice alternatives, and a green checkmark next to the correct alternative (i.e., immediate feedback was administered during quizzes). The experimenter read aloud the question stem and correct answer, and then moved on to the next question. Each clicker quiz was comprised of eight questions, and each quiz took approximately 7-9 minutes to complete.

At the beginning of each chapter unit, subjects completed a pre-quiz via clickers without having read the textbook chapter. Immediately after the pre-quiz (i.e., on the same day), subjects began reading the Social Studies chapter in their textbook and listened to a lecture about the chapter material. After 2-3 school days (during which the subjects completed reading the chapter and the classroom teacher covered all chapter material), subjects completed a post-quiz via clickers. Two days later, during which the classroom teacher reviewed all chapter material, subjects completed a review quiz via clickers.

Two days later (during which the classroom teacher reviewed material again), subjects completed a final chapter test. Final chapter tests were administered online (via Google Docs, http://docs.google.com), while subjects sat

individually at a PC computer. The chapter test was self-paced, and subjects viewed each multiple-choice question one at a time. Once students selected a multiple-choice alternative, they moved on to the next question; however, the online chapter test also allowed subjects to go back (question by question) if they wanted to change their answer. Once subjects responded to all 24 questions, subjects were no longer able to return to the test to change their answers. No feedback was provided during the final chapter test.

In sum, subjects participated in only three within-subject learning conditions, crossed with two delayed test types. Dependent variables collected included only accuracy on test questions. The entire procedure was followed for both chapters of material, and lasted approximately one hour across 7-8 school days for each chapter. At the end of the experiment, subjects were debriefed and thanked for their time.

Results

Thirty-four students were absent during at least one quiz or exam, and their data were excluded from the reported analyses to ensure the integrity of the procedure. Thus, data reported are from 88 students (*M* age = 11.58 years, 48 females). A very similar pattern of results was found when data from all students who assented to participate were included (i.e., *n* = 122 absent and present students, excluding special education students; see Appendix G).

Initial quiz performance. Initial quiz performance for the first (pre-quiz), second (post-quiz), and third (review) quizzes are displayed in Table 7. In general, performance increased from the pre-quiz (38%) to the post-quiz (71%)

and also to the review quiz (84%), due to textbook reading, classroom lectures, and immediate feedback received on quizzes. Across the initial quizzes, performance was slightly greater in the mixed quiz condition (3X, 66%) compared to the higher order quiz condition (3X, 62%), likely due to the inclusion of easier concept questions in the mixed quiz condition.

A 2 [learning condition: higher order quizzes (3X), mixed quizzes (3X)] x 3 (quiz type: pre, post, review) repeated measures ANOVA on initial quiz performance revealed a marginal main effect of learning condition, $F(1, 87) = 3.55$, $p = .063$, $\eta_p^2 = .039$, and a significant main effect of quiz type, $F(2, 174) = 442.05$, $p < .001$, $\eta_p^2 = .84$; however, the interaction was not significant, $F(2, 174) = 2.03$, $p > .05$. In other words, students' initial quiz performance increased across the three quizzes, and did so similarly for both the mixed quiz and the higher order quiz conditions.

Final chapter test performance. Performance on the final chapter tests is displayed in Table 7 and Figure 4. Delayed test performance (collapsed over test type) was greatest for the mixed quiz (3X) condition (86%), followed by the higher order quiz (3X, 70%) and non-quizzed (60%) conditions. In addition, delayed test performance was similar for the final concept (73%) and final higher order (71%) tests. A 3 [learning condition: higher order quizzes (3X), mixed quizzes (3X), non-quizzed] x 2 (delayed test type: concept, higher order) repeated measures ANOVA on final test performance revealed a significant main effect of learning condition, $F(2, 174) = 128.98$, $p < .001$, $\eta_p^2 = .60$, a marginal main effect of delayed test type, $F(1, 87) = 3.19$, $p = .078$, $\eta_p^2 = .04$, and a

significant interaction between learning condition and delayed test type, $F(2, 174) = 28.30$, $p < .001$, $\eta_p^2 = .25$.

Regarding delayed performance on the final concept test, the mixed quiz (3X) condition resulted in far greater performance (91%) than the higher order quiz (3X) and non-quizzed conditions (64% each), $t(47) = 12.24$, $p < .001$, $d = 1.44$ and $t(47) = 13.63$, $p < .001$, $d = 1.55$, respectively. Consistent with Experiments 1 and 2, retrieval practice with higher order quizzes did not enhance delayed concept performance, but the mixed quiz (3X) condition (which included concept questions) produced a significant retrieval practice effect on the delayed concept test.

For delayed performance on the final higher order test, the mixed quiz (3X) condition resulted in marginally greater performance (82%) compared to the higher order quiz (3X) condition (75%), $t(87) = 2.27$, $p = .078$ ($p = .026$ without Bonferroni correction), $d = 0.34$, and significantly greater performance compared to the non-quizzed (56%) condition, $t(87) = 12.24$, $p < .001$, $d = 1.37$. In contrast to Experiment 2, mixed quizzing produced the greatest level of performance on both concept and higher order delayed tests, while providing a marginal benefit above and beyond the benefit received from higher order quizzing on delayed higher order performance.

Finally, the higher order quiz (3X) condition produced significantly greater final higher order test performance compared to the non-quizzed condition, $t(87) = 7.87$, $p < .001$, $d = 0.99$, replicating the finding from Experiments 1 that retrieval

practice with higher order questions improved delayed higher order performance compared to no quizzing.

Discussion

In Experiment 3, retrieval practice with three higher order quizzes improved delayed higher order performance by 19% compared to no quizzing, but higher order quizzing did not enhanced delayed concept performance. Critically, the mixed quiz condition (in which subjects received three quizzes comprised of both concept and higher order questions) improved both delayed concept and delayed higher order performance by 27% and 26%, respectively (compared to no quizzing). In addition, mixed quizzing also produced an additional benefit compared to higher order quizzing for delayed higher order performance by 7%, although this difference was marginally significant.

While benefits from retrieval practice in the present experiment are compared to a no-quiz condition (in which subjects studied all material, but did not receive quizzes), previous research has demonstrated similar benefits from retrieval practice over and above restudying with the same population of students, even over a long delay (see Roediger et al., in press). Even so, it is probable that the benefits from retrieval practice found in the present experiment would be lessened if compared to a restudy exposure control.

In sum, Experiment 3 replicated previous research in that retrieval practice improved delayed performance (compared to no quizzing) in an applied setting (see McDaniel, Agarwal et al., 2011; McDaniel, Thomas et al., 2011; Roediger et al., in press). In addition, the findings in Experiment 3 provide the first set of

evidence that higher order quizzing can be used to improve higher order learning in middle school classrooms, while also replicating findings from Experiments 1 and 2 in that higher order quizzing is a potent strategy for enhancing long-term higher order skill performance.

In a departure from Experiment 2, mixed quizzing produced some improvement (7%) in delayed learning over and above higher order quizzing on delayed higher order performance. Consistent with the transfer appropriate processing framework, engaging in both types of processing during initial learning (in the mixed quiz condition) enhanced delayed performance, possibly due to the greatest overlap in processing to the two final test types. Procedural departures from the previous experiments (described in the General Discussion) may have also influenced the results in Experiment 3, although precise determination of why mixed quizzing was slightly more beneficial than higher order quizzing in this particular experiment awaits future research.

General Discussion

The development of higher order skills is a critical component of education, yet the relationship between fact or concept learning and higher order learning is relatively unknown. In this project, I examined whether retrieval practice, typically used to enhance fact learning, can be used as a strategy to improve higher order skills in both laboratory and applied settings. Specifically, the present experiments demonstrated that retrieval practice with higher order quizzes enhances performance on higher order questions after a delay. These results were obtained under laboratory conditions with college students

(Experiment 1, which included exposure controls, as well as Experiment 2) and also replicated under applied conditions at a local middle school (in Experiment 3).

Generally, the results from Experiment 1 confirmed that concept quizzing improved delayed concept test performance (by 24% compared to studying or restudying), and Experiment 1 also established that higher order quizzing improved delayed higher order test performance (by 23-28% compared to studying or restudying). In Experiment 2, the use of two concept or two higher order quizzes produced an additional benefit for delayed performance by approximately 10%. Mixed quizzing (in which college students were presented with a mix of concept and higher order quizzes) in Experiment 2 also improved both concept and higher order performance after a two-day delay, albeit to a lesser extent than receiving two quizzes of the same question type. Experiment 3, conducted with 6th grade students, provided a conceptual replication in that higher order quizzing again improved delayed higher order performance and mixed quizzing also produced a benefit on both delayed concept and higher order performance (compared to no quizzing). Critically, in Experiment 3, the benefit from mixed quizzing was *greater* than that from higher order quizzing (by 7% on delayed higher order test), although this difference was marginally significant. In general, however, both higher order quizzing and mixed quizzing can be used to improve delayed higher order learning, a novel finding not yet established in the retrieval practice literature.

In discussing the present findings, first I consider the discrepant findings between Experiments 2 and 3. Second, I consider some potential limitations of the current research. Finally, I consider the theoretical and educational implications of the present study.

Higher Order versus Mixed Quizzing for Long-Term Learning

The results from the present study provide the first set of empirical evidence that retrieval practice can be used as a strategy to improve middle school and college students' higher order skills. Previous research has established the use of retrieval practice for enhancing fact learning (e.g., McDaniel, Roediger, & McDermott, 2007; Roediger et al., 2010; Roediger & Karpicke, 2006a) and also for enhancing the transfer or classification of knowledge (e.g., Butler, 2010; Rohrer, Taylor, & Sholar, 2010; Jacoby et al., 2010). Even so, the present study used long passage materials (1,000 – 2,500 words) and multiple-choice quiz questions designed to engage critical thinking skills including application (e.g., transfer), analysis (e.g., differentiating), evaluation (e.g., critiquing), and creation (e.g., predicting outcomes in new situations; see Figure 1 for a revised Bloom's taxonomy). Concept questions in the present study also departed from previous research which used detailed fact questions, in that the present concept questions were designed to encompass key ideas from passages, rather than details such as names, dates, vocabulary words, definitions, and so on.

In the present study, two retrieval practice conditions improved higher order learning: retrieval practice with higher order questions and retrieval practice

with a mix of concept and higher order questions. In Experiment 2, higher order quizzing was more potent for answering delayed higher order test questions, whereas in Experiment 3, mixed quizzing was more potent (albeit marginally) for answering delayed higher order test questions.

If the improvement of higher order learning is one's goal, should students and teachers be encouraged to use higher order or mixed quizzing? The results from Experiment 2 and from Experiment 3 suggest different recommendations: higher order quizzing in the former case, mixed quizzing in the latter case. As mentioned earlier, procedural changes from Experiment 2 to Experiment 3 may be responsible for the discrepant findings. I discuss a few possible explanations, although determination of the precise causal factor for the different findings awaits future research.

First, college students were included in Experiment 2, whereas middle school students were included in Experiment 3. Mixed quizzing may be more advantageous for students who have limited experience with higher order materials (i.e., 6th grade students) by strengthening conceptual representations (via concept questions in the present experiment) alongside retrieval practice with higher order questions followed by immediate feedback. In contrast, for college students the additional "scaffolding" from concept questions did not to transfer to an improvement in higher order performance. Instead, perhaps college students are more experienced at extracting conceptual information and simply benefit from more time spent on higher order retrieval practice. Unfortunately, mental effort ratings were not collected in Experiment 3 due to

classroom constraints; in future research, these ratings should be included in an applied setting, as they may address whether mixed quizzing reduced cognitive load (perhaps via scaffolding) for younger students.

Second, different retrieval practice conditions (e.g., quizzing with concept questions, restudying a passage, quizzing with higher order questions, etc.) were administered one after another during Session 1 in Experiment 2 for college students, whereas middle school students in Experiment 3 spent nearly two weeks with one type of retrieval practice (higher order or mixed) and then another two weeks with the other type of retrieval practice. It may be the case that when college students were exposed to all conditions at once, they expended more effort during higher order quizzes (and perhaps had more motivation to do well) because those quizzes were more challenging than concept or mixed quizzes by immediate comparison. Of course, this explanation is purely speculative, and awaits future research with college students (or laboratory research with middle school students), perhaps while comparing within-subject and between-subject manipulation of quizzing type (higher order or mixed).

Third, the mixing of quizzes across the two experiments diverged. In Experiment 2, subjects were first given a concept quiz (or higher order quiz) on a passage, next completed quizzes for other passages and conditions, and then received the other type of quiz (higher order or concept, respectively). In Experiment 3, on the other hand, each of three quizzes included a mix of question types: four concept and four higher order questions (although question

type was blocked during quizzes in Experiment 3). The spacing of question types over time in Experiment 2 versus the "massing" (i.e., immediate presentation) of question types within a quiz in Experiment 3 may have led to more potent benefits from mixed quizzing in Experiment 3. Again, this issue could be addressed in future research by specifically examining the massing or spacing of question types within or across quizzes.

A final potential account for the discrepant findings is that final test questions were rephrased from initial quiz questions in Experiment 2, whereas final test questions were identical to initial quiz questions in Experiment 3. Perhaps in Experiment 3, because final test questions were identical to initial quiz questions, a benefit from mixed quizzing emerged. Of course, Experiment 2 was carried out over the course of two days whereas Experiment 3 was carried out over the course of weeks, so the potential for "memorization" of identical questions (in 6th graders, nonetheless) over a longer time span in Experiment 3 may have been minimal.

In sum, it remains unclear why delayed higher order performance was most improved following higher order quizzing in Experiment 2, but slightly more improved following mixed quizzing in Experiment 3. Even so, both types of retrieval practice are recommended, although mixed quizzing may be more beneficial if the goal is to improve both long-term concept and higher order learning; however, bear in mind that the benefits from mixed quizzing may be reduced compared to the benefits from higher order quizzing on delayed higher order performance (as found in Experiment 2).

Potential Limitations

In addition to the procedural considerations discussed above, there are a few other potential criticisms of the current study. First, Experiments 2 and 3 did not include exposure control conditions. Previous laboratory and applied research has demonstrated that retrieval practice produces improvements in learning over and above exposure controls (e.g., restudying; see Carrier & Pashler, 1992; Jacoby et al., 2010; Roediger et al., in press; Roediger et al., 2006a), and Experiment 1 in the present study replicated these prior findings. While controlling for exposure is important, the principal aim for Experiments 2 and 3 was to examine the benefit of retrieval practice with higher order questions vs. mixed questions, not the benefit of retrieval practice over and above restudying. In addition, one can compare the results from Experiments 1 and 2 and see that mixed quizzing improved delayed performance in Experiment 2 more than restudying improved delayed performance in Experiment 1. Although the procedures from the two experiments were similar, this comparison is tentative due to the danger of cross-experiment comparisons. Still, considering the plethora of research demonstrating that benefits from retrieval practice are not due to exposure alone, and considering the results from Experiment 1, I argue that the inclusion of other types of quizzing and question types in lieu of exposure controls in Experiments 2 and 3 allowed for a valuable addition to the existing literature on retrieval practice.

Second, consider that Experiments 1 and 2 included slightly rephrased final test questions and Experiment 3 included final test questions identical to

initial quiz questions. Given the interest in examining higher order learning, future research should include greatly rephrased or even new questions (i.e., initially non-tested items) during the final test to ensure that subjects are engaging in higher order skill use on new concepts and questions, avoiding the potential for memorization of individual items (as may have been possible in the current study). Of course, if memorization were a major concern in any of the present experiments, then one may expect ceiling performance of 80-100% across conditions, which was not demonstrated – after a delay, performance remained at about a 50-70% accuracy level across all three experiments using multiple-choice questions and immediate feedback during initial learning.

Third, another potential criticism of the current study includes the use of multiple-choice questions, particularly when examining higher order learning. Some educators argue that higher order skills simply cannot be measured, at least not adequately, using the assessment techniques commonly found in classrooms, namely multiple-choice tests, closed-book tests, and intermittent evaluations (Ausubel et al., 1978; Martinez, 1999). Instead, educators advocate for paper assignments, essay tests, open-book tests, ongoing "portfolio" evaluations, and the teacher's judgment to determine higher order skill achievement (Hart, 1994; Kohn, 1999).

According to some educators, when questions or prompts are open-ended, students are able to describe their reasoning, explain a concept or theory, and compare ideas (Hart, 1994; Kohn, 1999). On the other hand, Bloom et al. (1956) noted that simply because students may be generating a response does

not imply that students are always using higher order skills to do so. Also, these assessment techniques are less objective precisely because there may be more than one correct response. The subjectivity of open-ended assessment techniques may undermine their utility as reliable tools to assess knowledge and higher order skills across students, situations, and time (Jordan, 1953). Furthermore, Jordan (1953) and Stanley and Hopkins (1972) explained that with careful construction of distracter items, the greatest advantage of multiple-choice questions is that they can be used to measure higher order skills by requiring students to distinguish carefully among the multiple alternatives (see also Haladyna, 1997).

In the interest of measuring learning objectively, higher order test questions were operationalized (see pp. 26-29) and multiple-choice items were carefully constructed to require subjects to apply conceptual knowledge, discriminate and evaluate arguments, and predict outcomes in new situations (see Appendices C, D, and F for test questions developed for the current study). It is possible that in using multiple-choice items, an element of higher order learning may have been lost. While future research could include the use of open-ended questions or even essay questions to examine higher order learning, I argue that the materials were carefully constructed and that the multiple-choice questions in the current study adequately engaged and measured students' higher order thinking and learning processes.

A final consideration is the hierarchical nature of data in Experiment 3 – middle school students were assigned to certain class periods of Social Studies

before the current study began, and this assignment was non-random. The subjects included in the two groups or sets used in Experiment 3 were also not assigned randomly, but rather, assigned by class period. Thus, the data from students is nested within class period, and again nested within set. The design in Experiment 3 was carried out completely within-subjects; however, nested individuals tend to be more alike than individuals selected at random (Raudenbush & Bryk, 2002). Unfortunately, because the number of levels within nests was low in Experiment 3, the use of a multilevel model to determine the influence of class period or set on performance was not possible. Future research should take care to examine the nested levels of data in applied settings, and apply multilevel or hierarchical models accordingly.

In sum, while these potential limitations (lack of exposure controls in Experiment 2 and 3, the use of rephrased/identical questions, the use of multiple-choice questions, and the hierarchical nature of data in Experiment 3) should be addressed in future research, I maintain that they do not negate the findings from the current study, namely that higher order and mixed retrieval practice can serve as potent strategies for improving delayed higher order performance. I turn, now, to discussing the theoretical implications of the present study.

Theoretical Implications

The present study makes four main contributions to the current literature on retrieval practice: 1) retrieval practice can be used to enhance higher order skill performance, 2) this finding holds in both laboratory and applied settings, 3) both higher order and mixed quizzing can be used as strategies to enhance long-

term higher order learning, and 4) enhanced concept learning (via retrieval practice) does *not* enhance immediate or delayed higher order performance. I consider these contributions in the context of the "thinking with the basics versus thinking is basic" debate discussed earlier, the transfer appropriate processing framework, and finally, cognitive load theory.

Recall that some people argue that thinking accompanies a foundation of basic knowledge, whereas others contend that thinking itself is natural and does not require a foundation of knowledge (e.g., Greeno, 1992). Can the results from the present study address this dispute? Well, yes and no. It is the case that retrieval practice with higher order quiz questions improved delayed higher order performance, supporting the notion that thinking skills can be improved "directly" without reinforcing basic understanding (e.g., of the passage materials in the current study), consistent with the "thinking is basic" viewpoint. At the same time, mixed retrieval practice with both concept and higher order quiz questions improved delayed higher order performance in one experiment (Experiment 3) but not in another (Experiment 2) when compared to benefits from higher order quizzing. In addition, quizzing with concept questions once or twice provided no benefit to delayed higher order performance, further evidence in contrast to the "thinking with the basics" viewpoint.

Thus, the current study provides some evidence in support of both sides of the debate, rather than differentiating the two arguments, but more strongly confirms the "thinking is basic" viewpoint in that higher order quizzing improved delayed higher order skill performance. It would be wise for future research to

tackle this theoretical and practical issue, particularly as educators who firmly believe in devoting classroom time to higher order thinking skills may be doing so at a detriment to both concept learning and also higher order skills, as well.

Regarding the transfer appropriate processing framework (Morris et al., 1977; see also McDaniel et al., 1978), all three experiments from the present study provide further evidence supporting the notion that a match in initial and final processing benefits long-term learning. Contrary to initial predictions under this framework, an improvement in higher order quizzing did not enhance delayed concept learning. Given the materials used in the current study, perhaps the higher order questions did not engage the type of processing required to answer the concept questions, preventing any transfer of processing from one question type to another.

Similarly, the mixed quiz condition in Experiments 2 and 3 enhanced both concept and higher order learning, most likely because the mixed quiz condition provided a match from initial concept and higher order processing to final concept and higher order processing. Consider, also, that the questions developed for the present study were not "yoked" between concept and higher order items. Perhaps a one-to-one relationship between question types would allow for appropriate and convergent processing to a greater extent than the current materials allowed.

Finally, according to cognitive load theory and the "thinking with the basics" viewpoint, it was predicted that concept quizzing would enhance delayed higher order learning, potentially because concept learning may reduce cognitive

load, thereby freeing up cognitive resources (e.g., working memory) and improving performance on higher order questions. Instead, results from the current study suggest that this is not the case – improved concept learning (via retrieval practice) did not produce a subsequent benefit to higher order learning on an immediate quiz (Experiment 2) or on delayed higher order final tests (Experiments 1 and 2).

Why *didn't* concept learning improve higher order skills, as many cognitive psychologists contend (e.g., Ausubel, 1961/1965; Bartlett, 1958; Bruner, 1959/1965, 1977; Hirsch, 1996; James, 1900; Willingham, 2009)? Emerging research suggests that retrieval practice enhances only those items that are similar across initial quizzes and final tests (Hinze & Wiley, 2011). While the simple explanation could be that fact or concept learning simply do not accompany higher order learning, I consider two additional explanations: first, the multiple-choice questions used the current study did not engage enough retrieval or generational processes to enhance the transfer of knowledge from concept questions to higher order questions (or vice versa); and second, that subjects were unaware of the relation of information and did not transfer their knowledge without explicit instructions to do so.

For the first account regarding question format, Hinze and Wiley (2011) found benefits from retrieval practice on novel items after a delay when initial learning occurred via free recall compared to fill-in-the-blank questions. Hinze and Wiley argued that fill-in-the-blank questions required subjects to retrieve only "surface memories," not broader concepts from passages. In the current study,

while the concept questions used the current study were specifically designed to engage conceptual processing (i.e., more broad than item-specific conceptual processing), perhaps the multiple-choice question format hampered the initial processing required in order for transfer from concept learning to higher order questions to occur. In a similar vein, Butler (2010) found a benefit of retrieval practice on novel transfer items using short answer (cued recall) questions during initial and final learning; however, Butler did not include different question formats, so it is unclear whether he would have found similar transfer benefits using multiple-choice or fill-in-the-blank questions. Thus, researchers are encouraged to investigate the extent to which question format influences transfer of factual or conceptual knowledge to novel questions (Hinze & Wiley), inferential questions (Butler), and also to different complexities of higher order questions (in the current study).

For the second account regarding explicit instructions (or the lack thereof in the current study), Chan et al. (2006, Experiment 3) found a benefit of retrieval practice on novel items when subjects were instructed to adopt a "broad retrieval strategy" during study, whereas subjects who were told to adopt a "narrow retrieval strategy" did not demonstrate a benefit of retrieval practice on related novel items. Similarly, Butler (2010, Experiment 3) found a benefit of retrieval practice on far transfer to novel items when subjects were explicitly told that the final test was related to information learned during the initial session (see also Chan, 2009). Furthermore, a classic study by Gick and Holyoak (1980; see also Bransford et al., 1986) demonstrated that students' conceptual knowledge

remains “inert” when not explicitly told to use previously learned information on novel items. It may be the case, then, that students in the current study would have transferred their factual/conceptual knowledge to the higher order questions if explicitly instructed to do so; of course, this speculation awaits future research.

Educational Implications

The main purpose of this study was to examine whether retrieval practice can be used as a technique to improve higher order skills. Indeed, all three experiments confirmed that retrieval practice with higher order questions, or with a mix of concept and higher order questions, can be used as a strategy to improve long-term higher order thinking skills for both middle school and college students. Considering that research on retrieval practice has been mostly restricted to fact-based materials (McDaniel et al., 2007; Roediger et al., 2010; Roediger & Karpicke, 2006a), the present research may address teachers’ hesitation to use such a strategy in their classroom in lieu of more “active” types of strategies that are perceived to improve higher order learning (e.g., concept mapping; Karpicke & Blunt, 2011). Of course, more research needs to be conducted to examine whether other cognitive strategies that have been demonstrated to improve fact learning also improve higher order learning (e.g., spaced study, interleaved practice, mnemonics, etc.; for reviews, see Agarwal, 2011; Rohrer & Pashler, 2010). As a starting point, by demonstrating that retrieval practice improves both concept learning and higher order skills, teachers may be more willing to use to adopt this strategy in their classrooms.

References

Agarwal, P. K. (2011). *Cognitive strategies improve students' higher order skills*. Manuscript submitted for publication.

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*, 861-876.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (abridged ed.). New York, NY: Addison Wesley Longman, Inc.

Ausubel, D. P. (1965). In defense of verbal learning. In R. Anderson & D. Ausubel (Eds.), *Readings in the psychology of cognition* (pp. 87-102). New York, NY: Holt, Rinehart, & Winston. (Reprinted from *Educational Theory*, *11*, 15-25, 1961).

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York, NY: Holt, Rinehart, and Winston.

Banks, J. A., Beyer, B. K., Contreras, G., Craven, J., Ladson-Billings, G., McFarland, M. A., & Parker, W. C. (1997). *World: Adventures in time and place*. New York, NY: Macmillan/McGraw-Hill.

Bartlett, F. C. (1958). *Thinking: An experimental and social study*. Westport, CT: Greenwood Press.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of

human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). New York, NY: David McKay Company, Inc.

Bransford, J. D., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist*, *41*, 1078-1089.

Bruner, J. S. (1977). *The process of education*. Cambridge, MA: Harvard University Press.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181-202). New York: Cambridge University Press.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118-1133.

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*, 30-41.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633-642.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of

related material. *Journal of Experimental Psychology: General*, *135*, 553-571.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*, 153-170.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*, 42-45.

Cuban, L. (1984). Policy and research dilemmas in the teaching of reasoning: Unplanned designs. *Review of Educational Research*, *54*, 655-681.

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*, 223-234.

Dewey, J. (1944). *Democracy and education: An introduction to the philosophy of education*. New York, NY: The Free Press. (Original work published 1916).

Finsterbusch, K., & McKenna, G. (Eds.). (1984). *Taking sides: Clashing views on controversial social issues* (3rd ed.). Guilford, CT: Dushkin Publishing Group.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*, 213-216.

Gatto, J. T. (2011, January 26). Does test-taking help students learn? [Letter to the editor]. *The New York Times*, p. A24.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306-355.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Boston, MA: Allyn & Bacon.

Hart, D. (1994). *Authentic assessment: A handbook for educators*. Menlo Park, CA: Addison-Wesley Publishing Company.

Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, *19*, 290-304.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1441-1451.

James, W. (1900). *Talks to teachers on psychology: And to students on some of life's ideals*. New York, NY: Henry Holt and Company.

Jordan, A. M. (1953). *Measurement in education: An introduction*. New York, NY: McGraw-Hill Book Company, Inc.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528-558.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772-775.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151-162.

Kester, L., Paas, F., & van Merriënboer, J. J. G. (2010). Instructional control of cognitive load in the design of complex learning environments. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 109-130). New York: Cambridge University Press.

Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional classrooms and "tougher standards."* Boston, MA: Houghton Mifflin Company.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction?" *Psychological Science*, *19*, 585-592.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, *34*, 207-218.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43-52.

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 131-

152). New York: Cambridge University Press.

McDaniel, M. A., Friedman, A., & Bourne, L. E. (1978). Remembering the levels of information in words. *Memory & Cognition*, *6*, 156-164.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399-414.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*, 200-206.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2011). *Quizzing promotes transfer of target principles in middle school science: Benefits on summative exams*. Manuscript submitted for publication.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*, 61-64.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519-533.

Münsterberg, H. (1909). *Psychology and the teacher*. New York, NY: D. Appleton and Company.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-

solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429-434.

Paas, F. G. W. C., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1-4.

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, *86*, 122-133.

Paas, F. G. W. C., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*, 419-430.

Pearson, K. (1911). On a correction needful in the case of the correlation ratio. *Biometrika*, *8*, 254-256.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, *64*, 916-924.

Plaas, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive load theory*. New York: Cambridge University Press.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437-447.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA:

Sage Publications.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist*, *38*, 15-22.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 91-108). New York: Cambridge University Press.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, *43*, 223-225.

Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New frontiers in applied memory* (pp. 13-49). Brighton, U.K.: Psychology Press.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (in press). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181-210.

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249-255.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, *39*, 406-412.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209-1224.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 233-239.

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). *E-Prime 2 user's guide*. Pittsburgh, PA: Psychology Software Tools.

Stanley, J. C., & Hopkins, K. D. (1972). *Educational and psychological measurement and evaluation* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Sternberg, R. J., Grigorenko, E. L., & Zhang, L. (2008). Styles of learning and thinking matter in instruction and assessment. *Perspectives on Psychological Science*, *3*, 486-506.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295-312.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29-47). New York: Cambridge University Press.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837-848.

Tuovinen, J. E., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science*, *32*, 133-152.

Ward, D. (2007). eInstruction: Classroom Performance System [computer software]. Texas: eInstruction Corporation.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571-80.

Willingham, D. T. (2009). *Why don't students like school: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco, CA: Jossey-Bass.

Table 1

Initial Quiz and Delayed Test Performance (Proportion Correct) as a Function of Learning Condition From Experiment 1

	Initial Quiz	Final Concept Test	Final Higher Order Test	Delayed Average
Study Once		.54 (.21)	.44 (.18)	.49
Study Twice		.54 (.21)	.49 (.19)	.51
Concept Quiz (1X)	.59 (.17)	.78 (.19)	.46 (.22)	.62
Higher Order Quiz (1X)	.47 (.15)	.53 (.21)	.72 (.21)	.62
Average	.53	.60	.53	

Note. Standard deviations are displayed in parentheses.

Table 2

Initial Quiz and Delayed Test Reaction Time (Average Seconds Per Question Answered Correctly) as a Function of Learning Condition From Experiment 1

	Initial Quiz	Final Concept Test	Final Higher Order Test	Delayed Average
Study Once		16.79 (7.91)	21.44 (10.54)	19.12
Study Twice		17.79 (9.23)	20.69 (6.47)	19.24
Concept Quiz (1X)	16.47 (5.83)	12.91 (5.20)	21.58 (12.42)	17.24
Higher Order Quiz (1X)	21.14 (8.57)	15.90 (6.37)	15.05 (5.54)	15.47
Average	18.81	15.85	19.69	

Note. Standard deviations are displayed in parentheses.

Table 3

Initial Quiz and Delayed Test Mental Effort Ratings (Average Rating Per Question Answered Correctly) as a Function of Learning Condition From Experiment 1

	Initial Quiz	Final Concept Test	Final Higher Order Test	Delayed Average
Study Once		4.91 (1.51)	5.18 (1.66)	5.05
Study Twice		4.68 (1.46)	4.80 (1.35)	4.74
Concept Quiz (1X)	4.36 (1.14)	3.72 (1.21)	4.97 (1.58)	4.35
Higher Order Quiz (1X)	4.50 (1.04)	4.61 (1.40)	3.96 (1.30)	4.28
Average	4.43	4.48	4.73	

Note. Standard deviations are displayed in parentheses.

Table 4

Initial Quiz and Delayed Test Performance (Proportion Correct) as a Function of Learning Condition From Experiment 2

	Initial Quiz #1	Initial Quiz #2	Final Concept Test	Final Higher Order Test	Delayed Average
Higher Order Quiz (1X)	.47 (.11)		.54 (.23)	.77 (.17)	.65
Higher Order Quizzes (2X)	.49 (.14)	.83 (.12)	.53 (.22)	.85 (.13)	.69
Concept Quizzes (2X)	.57 (.17)	.91 (.08)	.90 (.13)	.48 (.19)	.69
Mixed Quizzes (2X)	.52 (.19)	.53 (.15)	.78 (.18)	.71 (.18)	.75
Mixed: Concept-Higher	.58 (.22)	.47 (.11)	.81 (.17)	.71 (.18)	.76
Mixed: Higher-Concept	.45 (.13)	.60 (.15)	.76 (.18)	.71 (.19)	.73
Average	.53	.76	.69	.70	

Note. Standard deviations are displayed in parentheses.

Table 5

Initial Quiz and Delayed Test Reaction Time (Average Seconds Per Question Answered Correctly) as a Function of Learning Condition From Experiment 2

	Initial Quiz #1	Initial Quiz #2	Final Concept Test	Final Higher Order Test	Delayed Average
Higher Order Quiz (1X)	19.83 (7.25)		15.14 (6.55)	12.75 (3.99)	13.94
Higher Order Quizzes (2X)	20.36 (8.00)	9.35 (2.56)	14.63 (5.13)	11.16 (3.72)	12.90
Concept Quizzes (2X)	16.47 (5.81)	8.20 (2.63)	9.86 (3.41)	17.92 (6.48)	13.89
Mixed Quizzes (2X)	18.26 (7.53)	16.27 (5.41)	11.17 (3.72)	13.59 (4.34)	12.38
Mixed: Concept-Higher	14.98 (6.34)	17.89 (6.42)	11.33 (2.98)	12.67 (4.24)	12.00
Mixed: Higher-Concept	21.54 (7.29)	14.64 (3.62)	11.01 (4.40)	14.50 (4.33)	12.76
Average	18.73	11.27	12.70	13.85	

Note. Standard deviations are displayed in parentheses.

Table 6

Initial Quiz and Delayed Test Mental Effort Ratings (Average Rating Per Question Answered Correctly) as a Function of Learning Condition From Experiment 2

	Initial Quiz #1	Initial Quiz #2	Final Concept Test	Final Higher Order Test	Delayed Average
Higher Order Quiz (1X)	4.71 (1.06)		4.57 (1.45)	3.72 (1.29)	4.14
Higher Order Quizzes (2X)	4.63 (1.13)	2.56 (1.21)	4.66 (1.40)	3.13 (1.01)	3.89
Concept Quizzes (2X)	4.25 (1.11)	2.41 (1.17)	3.10 (1.22)	4.85 (1.23)	3.97
Mixed Quizzes (2X)	4.40 (1.32)	4.29 (1.17)	3.60 (1.22)	3.76 (1.20)	3.68
Mixed: Concept-Higher	3.91 (1.28)	4.22 (1.25)	3.65 (1.27)	3.63 (1.25)	3.64
Mixed: Higher-Concept	4.89 (1.18)	4.35 (1.11)	3.55 (1.19)	3.89 (1.56)	3.72
Average	4.43	3.08	3.98	3.86	

Note. Standard deviations are displayed in parentheses.

Table 7

Initial Quiz and Delayed Test Performance (Proportion Correct) as a Function of Learning Condition From Experiment 3

	Pre-Quiz	Post-Quiz	Review Quiz	Final Concept Test	Final Higher Order Test	Delayed Average
Non-Quizzed				.64 (.18)	.56 (.18)	.60
Higher Order Quizzes (3X)	.38 (.16)	.68 (.21)	.82 (.17)	.64 (.20)	.75 (.21)	.70
Mixed Quizzes (3X)	.38 (.18)	.73 (.19)	.87 (.15)	.91 (.17)	.82 (.21)	.86
Average	.38	.71	.84	.73	.71	

Note. Standard deviations are displayed in parentheses.

Figure 1

An Illustration of the Revised Bloom's Taxonomy, Adapted from Anderson et al. (2001)

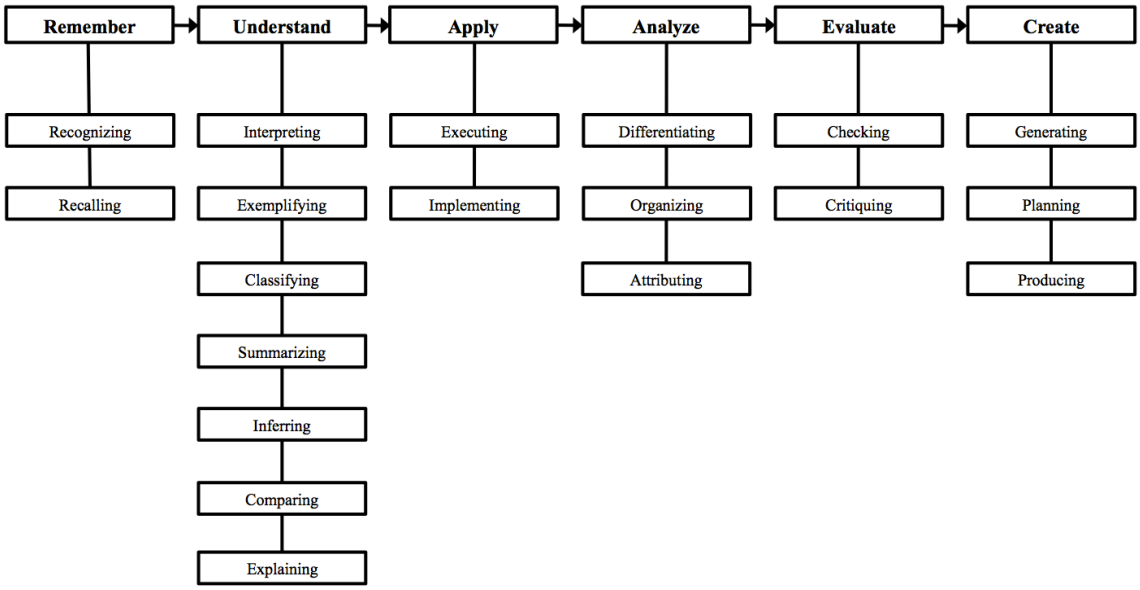
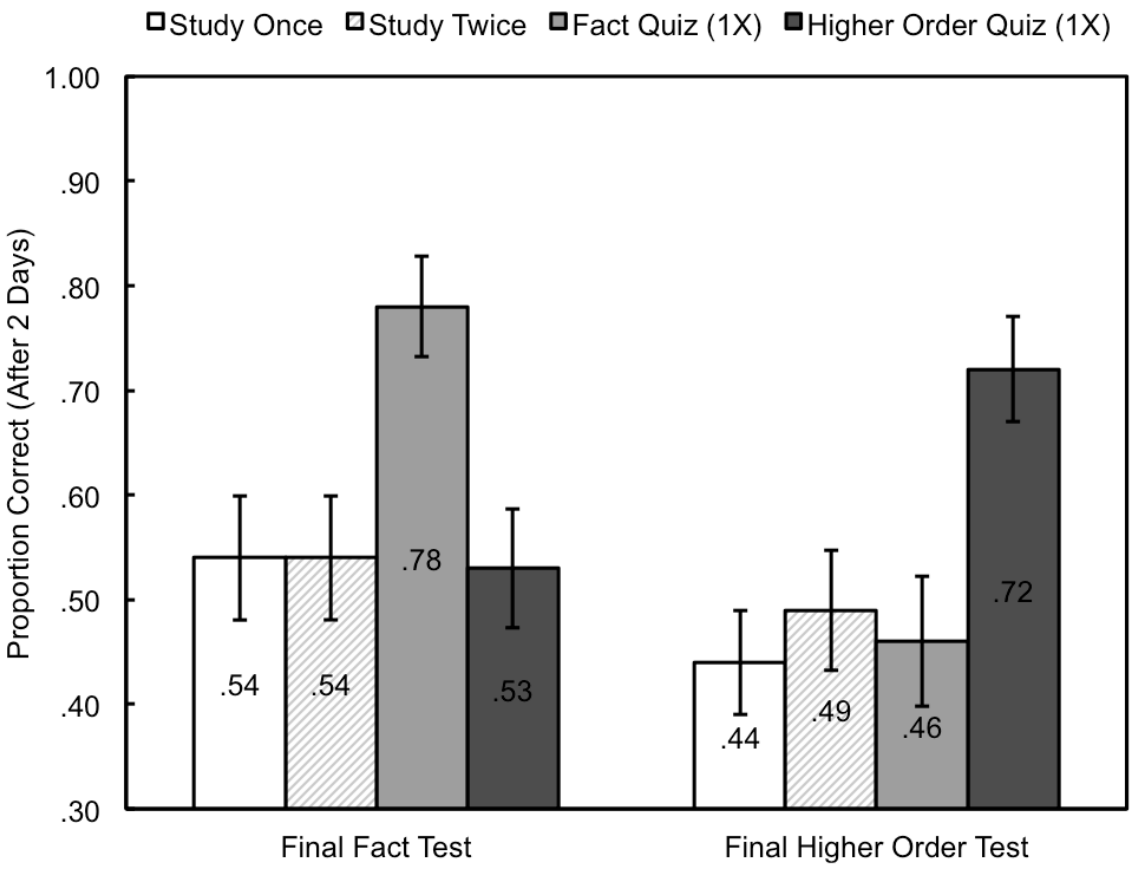


Figure 2

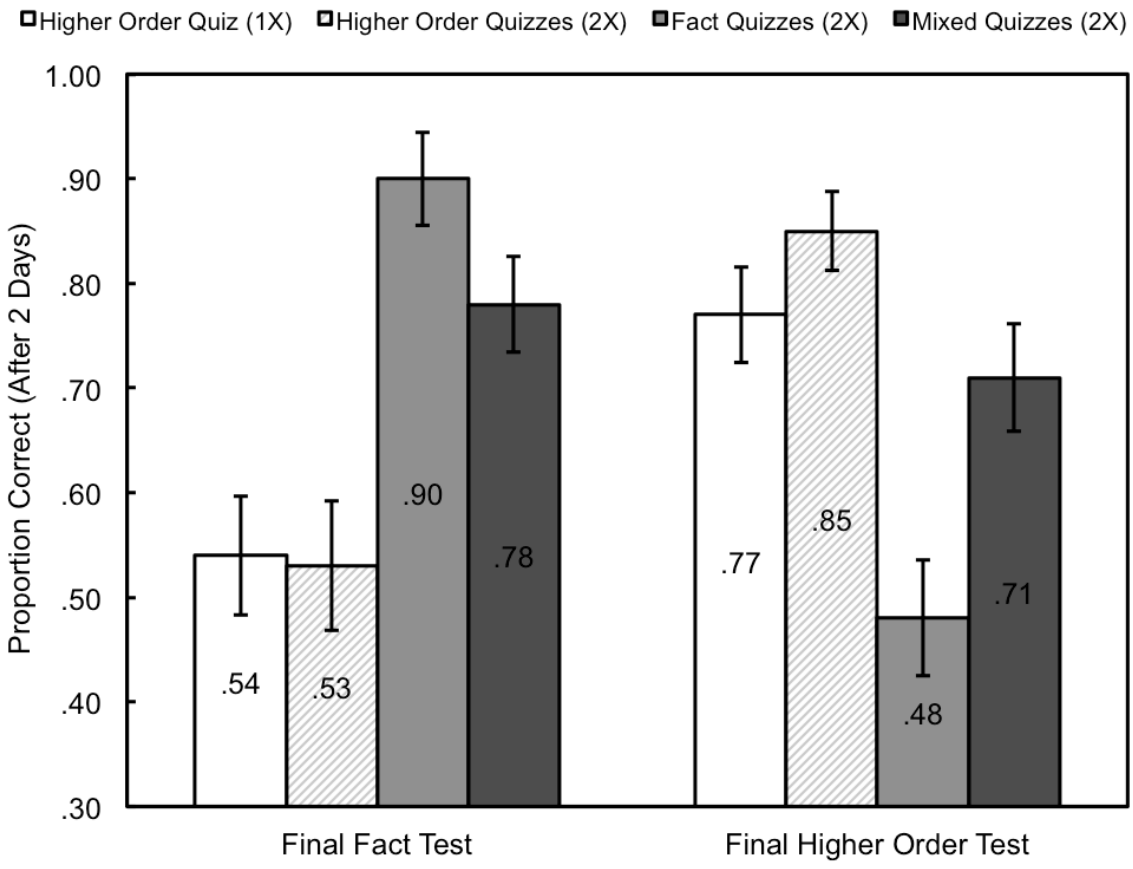
Delayed Test Performance (Proportion Correct After Two Days) as a Function of Learning Condition From Experiment 1



Note. Errors bars represent 95% confidence intervals.

Figure 3

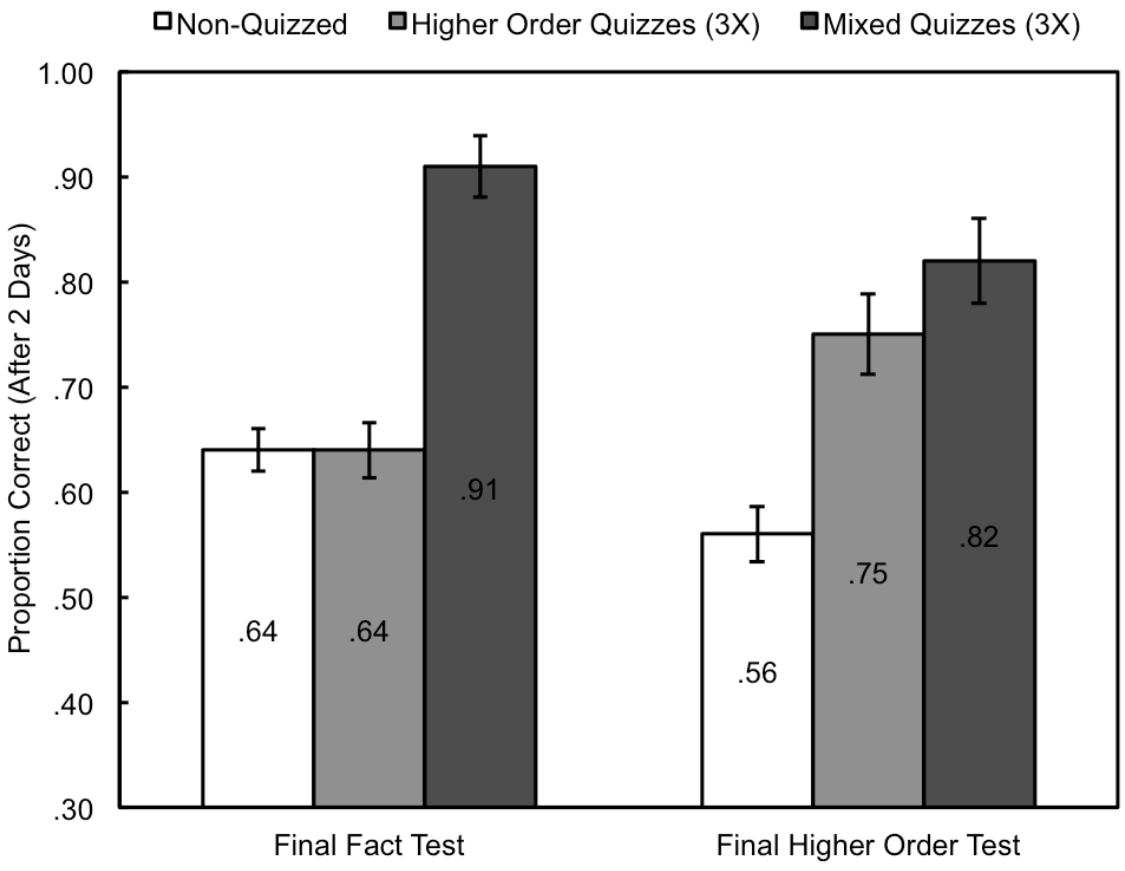
Delayed Test Performance (Proportion Correct After Two Days) as a Function of Learning Condition From Experiment 2



Note. Errors bars represent 95% confidence intervals.

Figure 4

Delayed Test Performance (Proportion Correct After Two Days) as a Function of Learning Condition From Experiment 3



Note. Errors bars represent 95% confidence intervals.

Appendix A

Counterbalancing Orders Used in Experiments 1, 2, and 3

Experiment 1

	Welfare	Vaccines	Multicul	Biotech	SexDiff	Lincoln	Superfund	WWII
1, 2	Study Twice	Study Twice	Concept	Concept	Study Once	Study Once	Higher Order	Higher Order
3, 4	Study Once	Study Once	Higher Order	Higher Order	Concept	Concept	Study Twice	Study Twice
5, 6	Higher Order	Higher Order	Study Once	Study Once	Study Twice	Study Twice	Concept	Concept
7, 8	Concept	Concept	Study Twice	Study Twice	Higher Order	Higher Order	Study Once	Study Once

Experiment 2

	Welfare	Vaccines	Multicul	Biotech	SexDiff	Lincoln	Superfund	WWII
1, 2	Concept 2X	Concept 2X	Mixed (H-F)	Mixed (H-F)	Higher 1X	Higher 1X	Higher 2X	Higher 2X
3, 4	Mixed (F-H)	Mixed (F-H)	Concept 2X	Concept 2X	Higher 2X	Higher 2X	Higher 1X	Higher 1X
5, 6	Higher 1X	Higher 1X	Higher 2X	Higher 2X	Mixed (H-F)	Mixed (H-F)	Concept 2X	Concept 2X
7, 8	Higher 2X	Higher 2X	Higher 1X	Higher 1X	Concept 2X	Concept 2X	Mixed (F-H)	Mixed (F-H)

Experiment 3

	Russian Revolution	World War II
Set A (three class sections)	Higher Order Only	Concept + Higher Order Mix
Set B (three class sections)	Concept + Higher Order Mix	Higher Order Only

Note. For Experiments 1 and 2, odd counterbalancing orders received final concept tests first during Session 2, alternating with final higher order tests. Even orders received final higher order tests first during Session 2, alternating with final concept tests.

Appendix B

Passages Used in Experiments 1 and 2

Passages and test questions used in Experiments 1 and 2 were adapted from books in the "Taking Sides" McGraw-Hill Contemporary Learning Series (http://www.mhcls.com).

WELFARE: Finsterbusch, K., & McKenna, G. (Eds.). (1984). *Taking sides: Clashing views on controversial social issues* (3rd ed.). Guilford, CT: Dushkin Publishing Group.

VACCINES: Daniel, E. L. (Ed.). (2006). *Taking sides: Clashing views in health and society* (7th ed.). Dubuque, IA: McGraw-Hill Companies, Inc.

MULTICULTURALISM: Noll, J. W. (Ed.). (2001). *Taking sides: Clashing views on controversial educational issues* (11th ed.). Guilford, CT: Dushkin/McGraw-Hill.

BIOTECH: Moseley, W. G. (Ed.). (2007). *Taking sides: Clashing views on African issues* (2nd ed.). Dubuque, IA: McGraw-Hill Companies, Inc.

SEX DIFFERENCES: Paul, E. L. (Ed.). (2002). *Taking sides: Clashing views on controversial issues in sex and gender* (2nd ed.). Guilford, CT: McGraw-Hill/Dushkin.

LINCOLN: Madaras, L., & SoRelle, J. M. (Eds.). (1993). *Taking sides: Clashing views on controversial issues in American history* (5th ed., Volume 1). Guilford, CT: Dushkin Publishing Group.

SUPERFUND: Easton, T. A. (Ed.). (2006). *Taking sides: Clashing views on environmental issues* (11th ed.). Dubuque, IA: McGraw-Hill Companies, Inc.

WORLD WAR II: Madaras, L., & SoRelle, J. M. (Eds.). (1993). *Taking sides: Clashing views on controversial issues in American history* (5th ed., Volume 2). Guilford, CT: Dushkin Publishing Group.

CONTRAS (passage presented during the instructional phase of Session 1): Rourke, J. T. (Ed.). (1987). *Taking sides: Clashing views on controversial issues in world politics* (1st ed.). Guilford, CT: Dushkin Publishing Group.

Does Welfare Do More Harm Than Good?

YES

New York is the most welfare-oriented community in the United States, and it is the most dramatic example of the results of trying to do good through government programs. Spending by the city government is larger relative to its population than in any other city in the U.S. But more money, more programs, and more taxes didn't work. They led to financial catastrophe without meeting the essential needs of the people. Let us take a closer look at a few other examples.

The major welfare-state program in the U.S. on the federal level is Social Security. On the one hand, it is a sacred cow that no politician can question. On the other hand, it is the target of complaints from all sides. Persons receiving payments complain that the sums are inadequate to maintain the standard of life they had been led to expect. Persons paying Social Security taxes complain they are a heavy burden. Taxpayers complain that the unfunded obligations of the Social Security system total many trillions of dollars, and that not even the present high taxes will keep it solvent for long.

Government programs to provide housing and medicine have also expanded rapidly. Housing programs started with government construction of housing units for low-income families. More recently, "rent supplements," or government subsidization of rents for privately owned housing units, have been added. In addition, the federal government has provided medical care for the military and veterans. After the introduction of Medicare and Medicaid, government spending on health mounted rapidly and the government's share of total expenditures on medical care has almost doubled. In terms of the initial objective, these programs have been a conspicuous failure. The public housing units have frequently become slums and hotbeds of crime. The inevitable result from the medical programs has been sharp increases in the price of medical care and in the incomes of physicians and others engaged in rendering medical services.

Why have all these programs been so disappointing? Their objectives were surely humanitarian and noble. Why have they not been achieved? As welfare programs expanded, the numbers changed. Legislators vote to spend someone else's money. The voters who elect the legislators are in one sense voting to spend their own money on themselves, but not in any direct sense of spending. The connection between the taxes any individual pays and the spending he votes for is exceedingly loose. In practice, voters, like legislators, are inclined to regard someone else as paying for the programs the legislator votes for. Bureaucrats who administer the programs are also spending someone else's money. Little wonder that the amount spent explodes.

Most of the present welfare programs should never have been enacted. If they had not been, many of the people now dependent on them would have become self-reliant individuals instead of wards of the state. In the short run, that might have appeared cruel for some, leaving them no option to low-paying, unattractive work. But in the long run, it would have been far more humane.

NO

The true social role of government is very wide and it penetrates into the remotest corners of our daily lives in ways that are so familiar to us we are scarcely conscious of them. In view of this, the currently popular call to "get government off our backs" seems rather ludicrous. Equally nonsensical is the assertion that the taxing power of the federal government should never be used to promote social change. However, the federal government is in the business of influencing social change every minute of every day. To eliminate its social role, its responsibility to promote constructive social change, would be to eliminate a vast part of its general role and would take us back to the earliest days of the Republic when we tried, unsuccessfully, to govern ourselves through a loose confederation of the states.

A sharp reduction of the social role of the federal government is not in the interests of the nation. Looking backward over the past few decades, we can see that it is myth, not fact, that federal social programs for the most part failed. On the contrary, social programs greatly reduced poverty, hunger, malnutrition, infectious disease, and infant mortality. They made health care much more widely available. They gave dignity and opportunity to many of our fellow citizens. In these and other ways, social programs accomplish a great deal. Why, we may ask, should we abandon a public policy approach that achieved so much?

If one believes that the development of people – all people, whatever their economic status, physical or mental characteristics, sex, or color – is our highest priority, because it is fundamental to economic growth and to national security, and if one believes that equity among individual Americans on a national basis is the cornerstone of a workable society, then one must favor strong participation by the federal government in meeting the nation's social needs. If, on the other hand, one is not particularly concerned about the prospect of social unrest ahead, if one does not fear the consequences of reduced investment in people for economic growth and national security, if equity on a national basis is not high on the agenda, and if one believes that the workings of a free market economy can take care of most of the nation's social ills, then there will be little desire to see the federal social role maintained.

Through a wise and skillful exercise of federal executive and legislative power, we have the ability to ensure that every American has a chance to reach his or her true potential; that discrimination against any person on account of race, sex,

or cultural background is eradicated; that the hungry are fed and the handicapped cared for; that every family has a decent place to live; that minimum standards of health care are available to all; and that the elderly are protected. All of this we can do, and we can do it with the resources that will be available to the nation, without sacrificing either our security or economic growth.

Should Parents Be Allowed to Opt Out of Vaccinating Their Children?

YES

Parents do not want their children to be injured or die from a disease or a vaccination. As guardians of their children until those children are old enough to make life-and-death decisions for themselves, parents take very seriously the responsibility of making informed vaccination decisions for the children they love. That responsibility includes becoming educated about the relative risks of diseases when compared to the vaccines aimed at preventing them.

Like every encounter with a viral or bacterial infection, every vaccine containing lab-altered viruses or bacteria has an inherent ability to cause injury or even death. But because so little medical research has been conducted on vaccine side effects, no tests have been developed to identify and screen out vulnerable children. As a result, public health officials have taken a “one size fits all” approach and have aggressively implemented mandatory vaccination laws, while dismissing children who are injured or die after vaccination as unfortunate but necessary sacrifices “for the greater good.” This utilitarian rationale is of little comfort to the growing number of mothers and fathers who watch their once-healthy, bright children get vaccinated and then suddenly descend into epilepsy, learning and behavior disorders, autism, diabetes, arthritis, and asthma. Some adverse reactions are fatal.

The unanswered question is: To what extent has the administration of multiple doses of multiple vaccines in early childhood, when the body’s brain and immune system is developing at its most rapid rate, been a cofactor in epidemics of chronic disease? The assumption that mass vaccination policies have played no role is as unscientific and dangerous as the assumption that an individual child’s health problems following vaccination are only coincidentally related to the vaccine.

Questions about vaccination can only be answered by scientific research into the biological mechanism of vaccine injury and death, so that profiles can be developed to distinguish between vaccine-induced health problems and those that are not. Whether the gaps in scientific knowledge about vaccines will be filled in this decade or whether they will remain unanswered in the next decade

depends upon the funding and research priorities set by Congress and the health industry.

All diseases and all vaccines are not the same, and neither are children. Parents understand the qualitative difference between options. They are calling for enlightened, humane implementation of state vaccination laws, including protections and exemptions for religious or conscientious beliefs. This is especially critical for parents with reason to believe that their child may be at high risk for dying or being injured by one or more vaccines but cannot find a doctor to write an exemption.

Parents, who know and love their children better than anyone else, have the right to make informed, voluntary vaccination decisions for their children without facing state-sanctioned punishment. Whether a child is hurt by a vaccine or a disease, it is the mother and father – not the pediatrician, vaccine maker, or public health official – who will bear the lifelong grief and burden of what happens to that child.

NO

If the U.S. population or any population regards immunizing children as optional, we risk having large numbers of children becoming vulnerable to the most deadly diseases known to man. Without immunizations, there would be a significant possibility that children would contract some of the diseases that are now waiting to come back. These include whooping cough, polio, measles, mumps, meningitis, and diphtheria.

It is important to understand the concept of public immunity vs. individual risk. Individual risk is always a possibility with any procedure, medication, new activity, or vaccine. The key to any program or new intervention is to minimize the risk. There is no question that vaccines are the safest, most risk-free type of medication ever developed. Nevertheless, occasionally children have been known to experience a bad reaction to a vaccine. It is not, however, good public policy to give those few at-risk situations priority over the goal of protecting the population as a whole from disease. If the pool of unimmunized children becomes large enough, then the disease may reemerge, possibly in epidemic proportions.

For example, there is no scientifically proven link between the measles vaccine and autism. It is assumed that there has been an increase in the diagnosis of autism because the definition for who would fall under that category has changed. In addition, parents and medical professionals are more aware of this condition and are more likely to pursue its diagnosis. Though there may be an increase in the number of children who are diagnosed with autism, there have been many studies completed that show that the measles vaccine does not cause autism.

Should parents be able to choose not to vaccinate their child without being barred from enrolling that child in school? Immunizing children is a public health issue. Public health laws in all 50 states require immunization of children as a condition of school enrollment. This is as it should be, since public health must take precedence. Immunizations have a clear community benefit and, therefore, individual preferences should not be permitted to expose the public to the hazards of infectious diseases.

It is clear that the risk of exposing children to infectious disease should there be a decline in immunizations is a risk to which the population of the U.S. should not be exposed. It is always regrettable when an individual case of an adverse event occurs no matter what might have taken place. These adverse events clearly affect the child and obviously the family as well, and there indeed is always an outcry when this does occur. However, with all safe, proven interventions, an exception could always occur given a normal risk ratio.

Should Multiculturalism Permeate School Curriculum?

YES

It is by now a truism that our country's public schools are undergoing a dramatic shift that reflects the growing diversity of our population. Yet many educators and the schools in which they work seem no better prepared for this change than they were a decade ago. Most educators nationwide are white, middle class, monolingual English-speaking women and men who have had little direct experience with cultural, ethnic, linguistic, or other kinds of diversity, but they are teaching students who are phenomenally diverse in every way.

Contrary to what the pundits who oppose multicultural education might say, multicultural education is not about political correctness, sensitivity training, or ethnic cheerleading. It is primarily about social justice. Given the vastly unequal educational outcomes among students of different backgrounds, equalizing conditions for student learning needs to be at the core of a concern for diversity. A concern for social justice means looking critically at why and how our schools are unjust for some students. It means that we need to analyze school policies and practices that devalue the identities of some students while overvaluing others.

Schools inevitably reflect society, and the evidence that our society is becoming more unequal is growing every day. Inequality is a fact of life, but many educators refuse to believe or accept it, and they persist in blaming children, their families, their cultural and linguistic backgrounds, or laziness as the culprits. Once educators accept the fact that inequality is alive and thriving in our schools, they can proceed to do something about it. Until they do, little will change.

We can no longer afford to behave as if diversity were a dirty word. Every day, more research underscores the positive influence that cultural and linguistic diversity has on student learning. Yet we insist on erasing cultural and linguistic differences as if they were a burden rather than an asset. To become effective teachers of all students, educators must undergo a profound shift in their beliefs, attitudes, and values about difference.

Anybody who has walked into a classroom knows that teaching and learning are above all about relationships, and these relationships can have a profound impact on students' futures. But significant relationships with students are difficult to develop when teachers have little understanding of the students' families and communities. The identities of non-mainstream students frequently are dismissed by schools and teachers as immaterial to academic achievement. It is only when educators and schools accept and respect who their students are and what they know that they can begin to build positive connections with them.

Because most educators in the United States have not had the benefit of firsthand experiences with diversity, it is a frightening concept for many of them. If we think of teaching as a life-long journey of personal transformation, becoming a multicultural person is part of the journey. However we begin the journey, what we say about diversity is severely limited by our actions. Acknowledging and affirming diversity is to everyone's interest, including middle class white students. Given the tremendous diversity in our society, it makes eminent good sense to educate all our students to be comfortable with differences.

NO

What began during the early part of this century as a shift towards increased awareness of ethnic and minority contributions to American history has evolved into a pedagogy that makes diversity and difference the prime movers of the curriculum.
Although learning should be lifelong, schooling is a finite process. Inevitably, additions to the curriculum made in the name of diversity and inclusion render the necessity of displacement. A curriculum can contain just so much, and because education succeeds only when it includes prolonged and in-depth consideration of specific books, authors, ideas, and historical events, more in education often is less.

Multicultural education is undermined by two fatal flaws. The first is that the more the curriculum represents a multicultural test based upon "exposure to diversity," the more shallow and superficial learning becomes. By disavowing the difficult dilemma of choosing what comes out, multiculturalism ultimately reduces education to its shallowest possibilities – the mere glossing over of diverse subject matter – and renders the kind of understanding that comes from intensive, prolonged study of selected material impossible to attain.

Multiculturalism's second fatal flaw is that it necessarily precludes the single most important requirement for successful education: coherent means to a discernible end. By denying the existence of desirability of a distinctive American culture, thereby repudiating the need for public education to assist in the process of assimilation, multicultural education is both aimless and rudderless. Multicultural curricula meander to and fro, touching fleetingly upon cultural tidbits of theoretically limitless diverse groups.

Contrary to the assertions of proponents of multiculturalism that limitless pluralism enriches education, the de-emphasizing of specific core material and factual knowledge in high school resulted in what it inevitably must have: a plague of ignorance. Multiculturalism's subordination of facts and knowledge to critical thinking skills demonstrates its educational bankruptcy, for any critical opinion worthy of a passing grade must evolve out of knowledge and be grounded in objective facts.

As is inevitable with a multicultural curriculum, in order to make room for diverse additions, one must make equivalent subtractions. Omitted from one such multicultural curriculum were Robert E. Lee, Alexander Graham Bell, Thomas Edison, Albert Einstein, and the Wright brothers. Ultimately, students educated within the vague parameters of this multicultural curriculum will learn the hard truth: that any opinion about the birth of our nation without the knowledge of the First Continental Congress or of the Civil War without considering Robert E. Lee is not based on sufficient factual knowledge and, therefore, has little or no value in the marketplace of ideas.

Emphasis on multicultural diversity within the curriculum is not America's only choice. Educators should continue to explore other possibilities such as more diversity of schools and less diversity within schools. It will not be until the educational bankruptcy of multiculturalism is exposed fully that the deconstruction of American public education will be halted successfully.

Will Biotech Solve Africa's Food Problems?

YES

Few would disagree that the many claims and counterclaims concerning what biotechnology can or cannot do to solve Africa's food insecurity problem have mainly been made by non-Africans. Although opinions differ regarding the role biotechnology can play in African development, all must agree about the urgency to eradicate the perpetual cycle of hunger, malnutrition, and death in a world of plenty. Since farming is the most important source of income and sustenance for about 75% of the population of Sub-Saharan Africa, there is no doubt that

agricultural biotechnology can make very substantial contributions toward increasing food production by rural resource-poor farmers.

In villages, constraints to crop production include pests, diseases, weeds, low fertilizer inputs, poor roads to markets, etc. For some of these constraints, biotech is the most promising recourse to alleviate them. For example, recent research shows that a pest that hinders legume production in Africa can be controlled by applying biotech tools. It is conceivable that the millions of dollars being wasted each year by anti-biotech activists elsewhere could go a long way to help build badly needed capacity for biotech research in Africa. Also, biotech for Africa should mostly be done in Africa and mostly by Africans themselves. And yes, this is being realistic, and it can be done, if there is consensus and goodwill.

A good example of how biotech can reach rural farmers involves a special program where the composition of farmers includes male and female farmers, oxen owners, different age groups from different sub-villages, etc. This program ensures that farmers participate in the research as partners with scientists and other actors, and enables scientists to also utilize indigenous knowledge in research and development. This prevents “cut and paste” approaches that may be foreign market-driven and which tend to provide short-term, quick-fix solutions to unique problems faced by small scale farmers in Africa, who have developed their own unique crops, cropping, and farming systems that cannot be changed without their full and careful involvement. Participatory methods increase farmers’ inputs in the decision-making process as well as in the dissemination of research products through their involvement in field trials, famers’ field days, surveys, and farmer-to-farmer diffusion of information. Obviously, this is not the only way that research results from the laboratory arrive at farmers’ fields, but it illustrates the fact that applied biotech research can be targeted and tied to meet specific needs of rural farmers.

We live in a world that has become an increasingly interdependent “global village” due to advances in information and transportation technology. In this global village, millions have plenty of food to throw away, while millions of others die daily because they have nothing to eat. Although Africans are thankful for development and relief aid, they are uncomfortable about their condition of continuous dependence on handouts that come in many forms with no permanent solutions apparently in sight. Self-sufficiency initiatives is one step in the right direction that deserves support from all those who want to help African scientists and farmers to feed their own people.

NO

Although hunger is sorely persistent throughout much of the developing world, Africa is the only region where it is actually getting worse. In Latin America and Asia, the past two decades have seen a modest decline in malnourishment

among children. That helps explain why, sooner or later, almost any major agricultural development will have to justify itself in an African context.

A biotech fix would be costly for the farmer, would increase chemical use, would add no other benefits to the system, and in any case, does not yet even exist. On the other hand, fallow periods, when land is allowed to "go wild," help maintain long-term productivity by reducing weed and pest infestations, and by allowing soil nutrient levels to recover. Improved fallowing is extremely low-cost and confers all the benefits mentioned above. It's also readily accessible. In at least a rudimentary form, the technique is already being used by tens of thousands of farmers in eastern and southern Africa. It is projected that 50 million farmers will be using improved fallowing within the next five to ten years.

One of the most interesting features of the improved fallow system is that it allows for forms of research and development that farmers can do on their own. But if innovation is to contribute to the welfare of farming, it will have to extend beyond issues of yield. After all, many U.S. and European farmers have been teetering on the brink of economic extinction for years, and a substantial number have gone over it – even though they produce some of the highest yields in the world. In most developing countries, agriculture is still the predominant way of life, so the economic health of farming is a basic social issue. This is why the agricultural status quo is a dangerous absurdity. Corporations that sell farmers seed and pesticide are making tens of billions of dollars in sales each year, and those that distribute, process, and retail the harvests are making hundreds of billions. But farmers themselves are now members of the poorest, and ironically, the hungriest occupation on Earth.

Biotech farming can boost yield dramatically, but such improvements aren't going to bring prosperity to farmers. Doubling and tripling yields doesn't make much of a difference if you can't get your product to market. One non-profit has expanded their agenda to include a kind of farmer empowerment. They now coordinate seven farm cooperatives so that local growers can capture the marketing and distribution advantages that come with scale. Instead of each farmer buying their own delivery truck and setting up their own office, the farm cooperative can pool its resources for a much larger delivery truck and office. Money can go directly into the farmer's pocket – no middleman to pay, no bills for agrochemicals or expensive seeds. Foreigners don't arrive with some technology with highly dubious potential. Instead, we have a local response to a local problem. And the response worked, because the produce was beautiful and the farmer got paid.

Should We Continue to Study Sex Differences?

YES

The common description of empirical research as showing that sex-related differences are small, unusually unstable across studies, and inconsistent with gender stereotypes arose in part from a feminist commitment to gender similarity as a route to political equality. It also arose from piecemeal and inadequate interpretations of the relevant empirical research. These interpretations failed to place research on sex-related differences in the context of other psychological research and often implied that findings that were very ordinary (in terms of magnitude, consistency, etc.) were rather exceptional. Given the new understanding of empirical findings that is evolving, research psychologists should think more deeply about the purposes for which their research may be used. Is psychological research that compares the sexes beneficial or harmful? Does this research foster or hinder the social change that would increase gender equality?

The fear is often expressed in feminist writing that differences become deficiencies for women because women are an oppressed group. Anxiety about sex differences is especially strong to the extent that scientists favor biological explanations, because this approach might produce a portrayal of women as innately inferior to men. Yet, contemporary research that has systematically examined whether the traits and behaviors ascribed to women are regarded as inferior to those ascribed to men has not found evidence for this generalized unfavorable perception of women. This research has shown that the stereotype of women is more positive overall than the stereotype of men, at least in contemporary samples of U.S. and Canadian college students. The sex differences that scientists have documented do not tell a simple tale of female inferiority.

Social scientific knowledge of sex differences could enhance women's ability to understand the antecedents of inequality and to improve their status in society. Nonetheless, the aura of danger surrounds research on sex differences. Some critics urge psychologists to stop this dangerous work or at least censor it in various ways. Each researcher must, of course, weigh the potential costs and potential benefits. If enough research psychologists conclude that the costs outweigh the benefits, research comparing the sexes will recede once again because it is too politically relevant. However, the scientific work now possesses a momentum of its own, as more investigators become caught up in the sheer excitement of discovery and theory testing.

Contemporary psychology has produced a large amount of research revealing that behavior is sex differentiated to varying extents. The knowledge produced in this area of science can be beneficial both in helping women and men to understand their natures and their society, and in suggesting ways to enhance

gender equality. Yet there surely are dangers that the new research will be used in far less beneficial ways by the forces of society. Therefore, the stresses between gender politics and the science of gender are not going to disappear. Never before in the history of psychology has such a formidable body of scientific information encountered such a powerful political agenda. The results of this encounter should be instructive to all psychologists who believe that psychology should serve human welfare as it advances scientific understanding.

NO

The ideology of gender differences is ubiquitous in mainstream and minority United States cultures and has enormous significance for personal and social life. Our widely shared and strong beliefs about differences between women and men in interests, competencies, and roles are not benign or neutral, and their consequences are profound and continuous throughout the course of one's life. While the idea of difference is understood as a comparison of persons on some dimension, it also is embedded in a history in which one gender is valued over the other. Thus, the significance of gender difference ideology for social life results not only from the idea of difference, per se, but from the inextricable union of difference and inequality, in both the origin of a gender difference ideology, and in its operation in contemporary life.

"The study of gender differences in psychology has been nothing but a growth industry; it's here to stay." This assessment is chilling, since such study is intimately related to our culture's determined effort that gender differences be maintained. Cataloging gender differences serves a primarily political, not scientific, purpose as it rationalizes and perpetuates differences in power, and contributes to the continuation of separate spheres for women and men. A gender difference ideology, which has such destructive consequences, can be challenged through the painstaking work of social scientists who continue to present evidence of similarity between women and men with similar backgrounds, in similar positions and similar situations. Such data seriously challenge the easy and popular cliché that women are from Venus while men are from Mars.

An ideology of gender difference serves inequality and power differentials by limiting our vision and restricting our possibilities. In addition, a gender difference ideology is a source of personal confusion, stress, interpersonal difficulties, and social unease since our gender beliefs are often not reliable predictors of how individuals actually behave. Genders need not be understood through dichotomous opposition. Similarly, minority groups need not be understood in terms of how each differs from a majority norm but rather in terms of the historical, social, political, and economic forces that have influenced them.

It is because our construction of gender is inextricably tied to inequality that our study of gender must focus on the process and conditions that underlie this

inequality. The typical focus, the ways in which women and men are "different," does not really help us celebrate diversity. An informed appreciation of gender-related diversity requires that we understand the continuing relationships between inequality and gender categories, that we always examine gender in its cultural context, and that we recognize the full range of gender diversity.

We must insist that diversity, a term not much in vogue, refers to an appreciation of human possibilities, and not to a parade of socially constructed differences. The ways in which we vary needs to be understood as illustrating the potential of human organisms of both sexes for learning so that we can appreciate our commonalities as equal members of the human family.

Was Abraham Lincoln America's Greatest President?

YES

In the flames of civil war, Lincoln underwent seemingly endless crises that might have shattered a weaker man. Here he was – a President who lacked administrative experience, suffered from chronic depression, hated to fire inept subordinates and bungling generals – thrust into the center of a deadly conflict. Here he was, forced to make awesome decisions in a war that had no precedent in all American history, a war without constitutional or political guidelines for him to follow. At the same time, Lincoln had to live with the knowledge that he was the most unpopular President the Republic had known up to that time.

From all directions came cries that Lincoln was unfit to be President, that he was too inexperienced, too inept, too stupid and imbecilic, to reunite the country. Melancholy and inexperienced though he was, Lincoln managed nevertheless to see this huge and confusing conflict in a world dimension. He defined and fought it according to this core of unshakable convictions about America's experiment and historic mission in the progress of human liberty.

Nowhere was the struggle more evident than in the nagging problem of slavery. Recall that what guided Lincoln in the matter of emancipation was his commitment, not just to the Union, but to what it represented and symbolized. Here, as in all war-related issues, Lincoln's devotion to the war's central idea – to preserving a system that guaranteed to all the right of self-government – dictated his course of action.

In 1862, Lincoln called on Congress to adopt an emancipation amendment. In 1864, the Senate adopted it by a vote of 38 to 6, but it failed to muster the required two-thirds majority in the House. After that, Lincoln put tremendous pressure on the House to approve the amendment, using all his powers of persuasion and patronage to get it through. With the outcome much in doubt,

Lincoln and congressional Republicans participated in secret negotiations never made public – negotiations that allegedly involved patronage, a New Jersey railroad monopoly, and the release of rebels related to congressional Democrats – to bring wavering opponents into line.

In 1865, the House adopted the present Thirteenth Amendment by just three votes more than the required two-thirds majority. When ratified by the states, the amendment would end human bondage everywhere in America. Lincoln had come a long distance from the harassed political candidate, opposed to emancipation lest his political career be jeopardized, convinced that only the distant future could remove slavery from his troubled land. The Proclamation had indeed liberated Abraham Lincoln, enabling him to act more consistently with his moral convictions. He was, then, a warrior for the American dream, prepared to do whatever was necessary to save it short of abandoning the dream itself. Putting aside his own aversion to bloodshed and violence, Lincoln ended up pounding all his southern foes into submission. And he did so because that was the surest way he knew to shorten the conflict, end the killing, and salvage his American dream.

NO

Of course, nothing that we can identify as part of Lincoln's legacy belongs to him alone. In some respects, the Emancipator was carried along with the tides. The first and most obvious item in my bill of particulars for indictment concerns Lincoln's dishonesty and obfuscation with respect to the nation's future obligations to the Negro, slave, and free. Lincoln, in insisting that the Negro was included in the promise of the Declaration of Independence, seemed clearly to point toward a radical transformation of American society. But at the same time, he added certain modifications to this high doctrine: modifications required by those of his countrymen to whom he hoped to appeal. It was an essential ingredient of Lincoln's position that he make a success at being anti-Southern or anti-slavery without at the same time appearing to be significantly pro-Negro. Lincoln's commitment was precisely of the sort that the North was ready to make: passing legislation to restrict the flow of Negroes into the North, while exploiting black labor in a conquered South. Lincoln's double talk left the North with a durable tradition of self-congratulation.

The second heading in this "case against Lincoln" has to do with Lincoln's management of the commercial and business life of the part of the Republic under his authority. Military necessity provided an excuse, an umbrella of sanction, under which the essential nature of the changes being made in the relation of government to commerce could be concealed. The inflationary policy of rewarding the friends of the government sustained. The euphemism of our time calls this "income redistribution." But it was theft in 1864, and is theft today. As chief executive, Lincoln supported heavy taxes. The war was a legitimate explanation for these measures. Lincoln's participation in huge subsidies for

railroads and in other legislation granting economic favors is not so readily linked to “saving the Union.” All of his life, Lincoln was a friend of the big corporations. There can be no doubt of Lincoln’s responsibility for the depressing spectacle of greed concerning which so many loyal Northern men of the day spoke with sorrow, disappointment, and outrage.

A large part of the complaint against Lincoln has to do with his expansion of the powers of the presidency. Lincoln believed there were “no limits” to his powers if he exercised them in the name of preserving the Union. Lincoln began his tenure as a dictator when, without interference from Congress, he summoned militia, spent millions, suspended law, authorized recruiting, decreed a blockage, defied the Supreme Court, and pledged the nation’s credit. But in my opinion, the capstone of this case against Lincoln is what he had done to the language of American political discourse that makes it so difficult for us to reverse the ill effects of trends he set in motion with his executive power. I am chiefly referring to his habit of wrapping up his policy in the idiom of Holy Scripture, concealing within a Trojan horse the moral superiority of an agenda that would never have been approved if presented in any other form.

Is the Superfund Program Successfully Protecting the Environment from Hazardous Waste?

YES

Superfund, one of the main programs used by the Environmental Protection Agency (EPA) to clean up serious, often abandoned, hazardous waste sites, has been improved considerably in recent years. Notably, progress has been made in two important areas: the development of risk assessments that are scientifically valid yet flexible, and the development and implementation of better treatment strategies.

Before 1995, the EPA’s assessment of potential public health risks at Superfund sites often assumed future residential use at the site, however unrealistic that scenario might be. This assumption would often result in the need for costly soil and waste removal remedies necessary to protect against hypothetical risks, such as those to children playing in contaminated soil or drinking contaminated ground water, even at sites where future residential use was highly improbable. After 1995, revised land use guidelines provided a basis for selecting more realistic future use scenarios, with projected exposure patterns that may allow for less costly remedies.

Potentially responsible parties also complained that there was little room to tailor remedies to the magnitude of cancer risk at a site, and that the same costly remedies would be chosen for sites where the cancer risks may differ by several

orders of magnitude. However, the EPA has now established a risk-based hierarchy for remedy selection. For example, if cancer risks at a site exceed 1 in 1,000 people, then treatment or waste removal or both might be required. Sites that posed a lower lifetime cancer risk could be managed in other ways, such as by prohibiting the installation of drinking water wells, which likely would be far less expensive than intrusive remedies.

Revisions to land use guidelines also refined the EPA's evolving remedy-selection criteria. For example, these revisions require an explicit consideration of the short-term effectiveness of a remedy, including the health and safety risks to the public and to workers associated with remedy implementation. The EPA has learned by experience that ignoring implementation risks, such as those associated with vapor and dust emissions during the excavation of wastes, could lead to the selection of remedies that proved costly and created unacceptable risks.

Cleanup efforts in Superfund's early years were dominated by containment and excavation-and-disposal remedies. But over the years, cooperative work by government, industry, and academia have led to the development and implementation of improved treatment technologies. More recently, there has been a dramatic increase in the use of source control treatment. Two types of source control technologies that have been widely used are incineration and soil vapor extraction. Although the use of incineration decreased during the 1990s because of cost and other factors, soil vapor extraction remains a proven technology at Superfund sites.

In recent years, the rate at which waste sites are being added to the National Priorities List has been decreasing dramatically. From 1983-1991, the EPA placed an average of 135 sites on the list annually. The rate dropped to an average of 27 sites per year between 1992-2001. In 1988, most waste sites were in the investigation stage, and the Superfund program was widely criticized as being too much about studies and not enough about cleanup. Superfund is now a program predominantly focused on the design and construction of cleanup remedies.

NO

The prairie at Tar Creek, in the northeast corner of Oklahoma, is punctured with 480 open mine shafts and 30,000 drill holes. Little League fields have been built over an immense underground cavity that could collapse at any time. Acid mine waste flushes into drinking wells. When the water rises in Tar Creek, a neon-orange scum oozes onto the roadside. Wild onions are saturated with cadmium, which may explain why three different kidney dialysis centers have opened here to serve a population of only 30,000.

It wasn't supposed to be like this. In 1980, Congress passed the "Superfund law," one of the boldest environmental statutes in U.S. history. But today, Superfund is a program under siege, plagued by partisan politics, industry stonewalling, and bureaucratic inertia. According to the General Accounting Office (GAO), 25% of Americans still live within four miles of a Superfund site, many of them are fields saturated with cancer-causing chemicals and other toxins. The GAO reports that the program's budget fell 35% over the past decade. According to the EPA's inspector general, 29 projects in 17 states were underfunded last year. According to a U.S. Senator, the federal administration has "allowed these sites to rot where they are."

Tar Creek is a case in point. Two decades after it was targeted on the very first Superfund priority list, the site is worse off than ever. Early on, the government confined its effort to the polluted creek, without looking at chat piles (the powdery output of mills after ore is extracted from rock), soil, air quality, or the danger of sink holes. Was it a lack of knowledge of the danger, as EPA claims? Or industry influence, as environmentalists charge? Whatever the reason, federal attorneys settled with mining companies for pennies on the dollar. Now, after fruitless efforts to contain 28 billion gallons of acid mine water, contamination is spreading across a vast watershed. And although the EPA trucked out toxic dirt from about 2,000 homes and schools, Tar Creek's children still show elevated lead levels at six times the national average.

At Tar Creek, many residents have given up hope. Even the EPA, which has spent $107 million at the site, isn't sure if it can ever be repaired. "We don't have an off-the-shelf remedy," says an EPA Superfund official. "What do you do with the enormous chat piles? When does cleanup become impractical? We have limited resources." In a show of no confidence, the Oklahoma legislature passed a $5 million buyout for all families with children under age 6. The head of the Tar Creek Steering Committee, a group of buyout supporters, veers between cynicism and despair. "They think we're poor white trash," he says bitterly. "The votes here don't affect any federal election, so why bother? We've agitated till we can't agitate anymore." Meanwhile, at Tar Creek, the toxic dust keeps blowing in the wind.

Did World War II Liberate American Women?

YES

Men suspected that women would be changed by their wartime work experience, and their reactions ranged from cautious welcomes to offensive attacks. Feminists of the period often exhorted women to change, warning that otherwise they would become subjugated like the women of Nazi Germany. Even a moderate and "feminine" magazine sounded trumpets for change.

One of the striking themes in oral histories is the desire of women to test themselves, stretch themselves, prove themselves. Many women proudly proclaimed how they had "held their own with men." In retrospect, this is probably what laid the groundwork for the transformation of someone from a woman who was "just a mother" to a self-confident participant in the wider world.

For the first time, many of these former war workers spoke up and challenged the male prerogative to make the big decisions. The money they had earned and saved lent them moral authority, but it was the confidence they had developed that enabled them to exert that authority. Studies of changing power relationships in the family in the 1950s have suggested that working class wives who had worked in the past participated more in these kinds of decisions. The work process itself engendered feelings and attitudes in the women that had a lasting effect.

Of what broader significance, then, was the changed consciousness of women that resulted from their wartime experience? For one thing, it contributed to the tide of rising expectations of women. That tide, ultimately, led to the birth and growth of a social movement for women in the 1960s, just as the rising tide of expectations among blacks fueled the civil rights movement. Furthermore, we must remember that the generation of older, married women who were so deeply affected was that of the mothers of those who built the current women's movement. Even if the mothers' experience had little direct effect on their own daughters, it may have helped foster the development of a working class feminist consciousness among young women.

Oral histories have revealed the often private and subtle ways in which individual women were changed by their wartime experience. These individual changes were not merely fleeting. For it is the changes that individuals experience that both push for and support social transformation. The connection is not always immediate or clear. There is usually a lag, with ideas preceding practice. For example, despite a growing belief in egalitarian marriage over the past forty years, household responsibilities only now are beginning to be equalized.

The potential for social transformation was created by the wartime need for women workers. For a brief period, images of women were revised, employment opportunities were expanded, and public policy was enacted that created new services for women. These were necessary, but not sufficient conditions. Social values also had to change, including women's definitions of themselves. Women's wartime experience played a vital role in that process of redefinition – the reverberations of which are still being felt today.

NO

Like the depression, World War II brought new challenges and new disruptions to families. For many who looked forward to building stable and secure homes after the depression, the war put their hopes on hold. When thousands of men were called to war, their unquestionably manly responsibilities as soldiers took precedence over their roles as breadwinners. While the men vanished to foreign shores to fend off the enemy, the women were left to fend for themselves.

The war emergency required society to restructure itself and it opened the way for the emancipation of women on an unprecedented scale. The potential for gender equality now had a chance to reach fruition. In response to the needs of an expanding wartime economy, public policy shifted dramatically from barring women from jobs to recruiting them. Married women were not only tolerated in the paid labor force, they were actively encouraged to take “men’s jobs” as a patriotic duty to keep the war economy booming while the men went off to fight.

However, nearly all the “men’s jobs” filled by women went back to men when the war ended. Even during the war, both the popular literature and the politicians urged married women to return to their domestic duties and single women to relinquish their jobs and find husbands when the hostilities ceased. This advice reflected not only the affirmation of home and family, but the prevailing suspicion of women – especially unmarried women – who entered the world of men.

The employment of women during the war, then, created a deal of ambivalence. While encouraged to enter the paid labor force, women’s public presence gave rise to concerns about the long-term effects of the changes that were taking place while the men were overseas. These concerns were eased by viewing women’s jobs as temporary extensions of patriotism and domestic responsibilities that resulted from the emergency situation.

The vast changes in gender arrangements that some feared and others hoped for never fully materialized. Actually, the war underscored women’s tasks as homemakers, consumers, and mothers just as powerfully as it expanded their paid jobs. Few women took jobs that were previously held exclusively by men, and those who did earned less than men. Although women demonstrated their eagerness for nontraditional work and proved themselves competent, few were able to retain those jobs after the war. As a result, wartime ultimately reinforced the sex-segregation of the labor force.

And so the potential for a new model family, with two equal partners who shared breadwinning and homemaking tasks, never gained widespread support. In the long run, neither policymakers nor the creators of the popular culture encouraged that potential. Instead, they pointed to traditional gender roles as the best means for Americans to achieve the happiness and security they desired. Public policies and economic realities during the depression and the war limited the

options of both women and men, and reinforced traditional arrangements in the home. Even during the war, Americans were heading homeward toward gender-specific domestic roles.

Is the United States Justified in its Support of the Contras?

YES

For many years, Nicaragua was ruled by a series of right-wing dictators, but it was overthrown by a leftist guerrilla movement, the Sandinistas. At first, the U.S. government tried to have normal relations with the Sandinistas, but relations deteriorated rapidly. The U.S. accused the Sandinistas of suppressing promised democracy, of supporting leftist rebels in El Salvador, and of building a military force capable of threatening Nicaragua's neighbors. The U.S. began action against the Sandinistas that included supporting the "Contra" (against) rebels, who consisted of several loosely tied groups of rebels.

The Sandinistas have increasingly repressed freedoms in Nicaragua, and it is about time we ceased being fooled by Sandinista propaganda. It is about time we recognized that it is Nicaragua's aggression that is the source of the conflict in Central America. The principal target of Sandinista aggression has been El Salvador. Nicaragua has provided massive support to the Sandinistas seeking to overthrow El Salvador's government. That support has included training, command-and-control headquarters, and weapons, ammunition, and other vital supplies. Nicaragua has served as a sanctuary for the Sandinistas and headquarters for their political arm. Nicaragua has publicly identified itself with the goals and methods of the Salvadoran guerrillas. The evidence of this activity is real, varied, and massive. Sandinista commanders have, one after another, described in compelling detail the dependence of the Salvadoran guerrillas on Nicaraguan-supplied weapons and supplies, on safehaven in that country, on communications and command services from Nicaragua, and on training conducted in or facilitated by Nicaragua.

Also, there are the confessions of the Sandinistas themselves. They have, on several occasions, stated their capacity to halt the aid being provided to guerrillas in El Salvador. And yet, Nicaragua would have us, and the world, believe that none of this evidence exists. Nicaragua would like us, instead, to pitch all this evidence out the window and take its flat, unsupported word that "in truth, it is not engaged, and has not been engaged in, the provision of arms or other supplies" to the guerrillas in El Salvador. Nicaragua would have us disregard the tens of thousands of dead, the hundreds of millions of dollars in economic damage, the immense human misery it has imposed on El Salvador, and take its word that it has not attacked El Salvador.

I believe that continued U.S. support for the Contras is essential to induce the Sandinista regime to enter into meaningful negotiations. We have too often been faced with Sandinista promises that evaporate when the immediate tactical basis for their issuance has disappeared. The U.S. House of Representatives' approval of the request for further assistance for the Contras should give the Sandinistas good reason to negotiate seriously. Our support for the Contras is designed only to encourage the Sandinistas to participate seriously and in good faith. The question now is whether the Sandinistas truly want peace.

NO

In 1983, the U.S. House Intelligence Committee noted that assistance to the Contras was not working and would not work because the pressure represented by the Contras had the opposite effect than it was meant to create. It hardened rather than softened the resistance of the Sandinistas. It produced results exactly opposite to those aimed for by the United States. Negotiations failed. Still, the issue of providing more assistance to the Contras has been before the Congress ever since 1983. Today, the program of assistance for the Contras is just as unlikely to succeed as the program proposed in 1983. It differs only in its size, in the number of Contra fighters proposed to be armed, and the intensity of warfare that will likely result if it is approved.

The U.S. is still murky in its explanation of goals, yet it is understood that the Contra regime will not be sufficiently strong to overthrow the Sandinistas. The Contras will exert enough pressure only to force the Nicaraguan government to negotiate seriously with the Sandinistas. This approach ignores intelligence assessments that the Sandinistas are unlikely to agree to negotiations for the simple reason that they would threaten the very basic structure by which it controls Nicaragua. The result will not be a Sandinista willingness to change the undemocratic nature of the regime; the result will be further repression.

Furthermore, the improvements in the Nicaraguan military arsenal (helicopters, artillery, and mobility) make the prospects for future Contra successes dim. The Contras remain without a political infrastructure inside Nicaragua or a clear political message to give to the Nicaraguan people. The Contra regime is no more likely to defeat the Sandinista government than before. It is, in fact, less likely to do so. The U.S.'s policy of pressure has not worked and will not work in the future. It continues to be the assessment of the U.S. intelligence community that only U.S. forces could truly resolve the conflict in Nicaragua on a military basis.

I am deeply concerned that, as in the past, the Sandinista government is clearly moving down the path away from democracy and pluralism. I have no confidence that additional assistance to the Contras will produce the democracy the U.S. seeks to achieve in Nicaragua. Press censorship, repression of the church, and restriction on political activities will continue and perhaps increase. It

is an unfortunate fact that continued and increased military pressure by the Contras will not cause the Sandinistas to change their policies. Even with increased military activity in Nicaragua, it is unlikely that the flow of assistance to the Contras will improve the situation. The House Intelligence Committee's review of the situation in 1983 and the record of the Contras since that time leads me to believe that the U.S. policy of additional assistance to the Contras will not work. It will, in fact, be counterproductive. I do not make this decision lightly, for the problem represented by Nicaragua is a serious one. But, it is a problem not likely to be solved by aiding the Contras.

Appendix C

Concept and Higher Order Questions Used in Experiments 1 and 2, Session 1

Note. Correct answers are underlined, and the type of higher order question (apply, analyze, evaluate, or create) is indicated below for illustrative purposes, but correct answers and higher order types were not revealed to subjects during testing.

Does Welfare Do More Harm Than Good?

Concept Questions

Which is one of the solutions the "yes" author proposes?

1) Eliminate Medicaid, but keep all veteran benefits
2) Keep all welfare programs, but reduce spending within each program
3) Eliminate some welfare programs and increase spending for remaining programs
4) Eliminate all welfare programs in the United States

According to the "yes" author, what is one reason welfare programs are so expensive?

1) A great deal of staff are needed to administer the programs
2) There is little connection between taxpayers and legislators
3) Recipients are dependent and require a lot of assistance
4) There are too many recipients and not enough taxpayers

According to the "yes" author, what is one downside of the Social Security program?

1) Taxes raised are not enough to help the federal government
2) Taxes should not be required from younger adults to pay for older adults
3) Taxes raised are not enough to keep the program sustainable
4) Taxes should not be required because older adults can take care of themselves

Which is the primary reason the "yes" author is against welfare programs?

1) Welfare programs don't benefit recipients or taxpayers
2) Welfare programs create dependence for recipients
3) Welfare programs are too expensive for taxpayers
4) Welfare programs are not the government's responsibility

What is one benefit of welfare programs that the "no" author supports?

1) They eradicate discrimination
2) They help support local communities
3) They are affordable and feasible

4) They help everyone, not just recipients

According to the "no" author, what is the purpose of taxation?
1) To provide citizens a way to support their government
2) To provide citizens with services they can't pay for on their own
3) To provide the government a way to act on citizens' behalf
4) To provide the government with means to improve society

Which is the primary reason the "no" author supports welfare programs?
1) They create independence, not dependence
2) They improve, not hinder, economic growth
3) They are the government's responsibility
4) They are a good investment of taxpayer money

According to the "no" author, a free market system
1) Can address problems of discrimination
2) Is insufficient to provide equality for citizens
3) Is the only alternative to welfare programs
4) Helps make welfare programs even stronger

Higher Order Questions

APPLY: What type of society would the "yes" author expect if there were no welfare programs in the future?
1) A society in which all individuals are self-reliant and independent
2) A society in which there would be no role for the government
3) A society in which no one would be required to pay taxes
4) A society in which all individuals are treated equally

ANALYZE: Which author would agree with the following statement? "It is honorable for the government to help society."
1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "yes" author's views?
1) Welfare programs can never work, because they are always too expensive
2) Welfare programs are harmful, because they make bad situations even worse
3) Welfare programs waste taxpayer money on people who don't really need help

4) Welfare programs could work, but they rarely meet the needs of the people

CREATE: How do you predict the "yes" author would react if he or she became unemployed and needed welfare assistance?

1) The "yes" author might accept government assistance, but would seek help from local organizations first
2) The "yes" author would not accept government assistance, but would try to find a new job
3) The "yes" author might accept government assistance, but would try to find a new job first
4) The "yes" author would not accept government assistance, but would seek help from local organizations

APPLY: What type of global government role would the "no" author support?

1) Governments around the world are obligated to help poor countries
2) Governments around the world are obligated to help when asked
3) Governments around the world are obligated to help all countries
4) Governments around the world are obligated to help countries that reciprocate

ANALYZE: Which author would agree with the following statement? "Investing in people is good for economic growth."

1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "no" author's views?

1) The government's primary role is advancing equality
2) The government's primary role is advancing morality
3) The government's primary role is advancing security
4) The government's primary role is advancing liberty

CREATE: Which tax and spending structure do you predict the "no" author would support?

1) Equal taxation of all Americans; equal spending on all Americans
2) Higher taxes for the rich, lower taxes for the poor; more spending on the rich, less spending on the poor
3) Equal taxation of all Americans; less spending on the rich, more spending on the poor
4) Higher taxes for the rich, lower taxes for the poor; less spending on the rich, more spending on the poor

Should Parents Be Allowed to Opt Out of Vaccinating Their Children?

Concept Questions

According to the "yes" author, vaccination shouldn't be mandatory because
1) We can't screen out vulnerable children
2) Research has verified its ineffectiveness
3) Vaccines do more harm than good
4) Vaccines are prohibited by most religions

According to the "yes" author, parents have
1) A responsibility to decide on behalf of their children
2) The right to decide on behalf of their children
3) An obligation to decide on behalf of their children
4) The option to decide on behalf of their children

The "yes" author argues that we need more research in order to
1) Determine the effectiveness of vaccines
2) Determine the long-term effect of vaccines
3) Determine the side effects from vaccines
4) Determine the mechanism behind vaccines

Which is the primary reason the "yes" author believes that parents should be able to opt out of vaccination?
1) Vaccination costs outweigh the benefits
2) Vaccination practices lack solid research
3) Vaccination for all children is too simplistic
4) Vaccination has the potential to cause death

According to the "no" author, an increase in autism diagnoses is not a result of the measles vaccine, but a result of
1) Poor childhood nutrition or immunity
2) A lack of understanding of autism
3) A change in the diagnostic definition
4) Some being more at-risk than others

According to the "no" author, giving parents the option to opt out of vaccination
1) Will only lead to more and more parents opting out
2) Will increase, not decrease, danger to the population
3) Is malpractice, and against state and federal law
4) Is a decision for medical professionals, not politicians

The "no" author argues that vaccines may always carry some amount of risk, but that this risk
1) Is a possibility with any medical procedure
2) Is too small to be of concern to the community

3) Should be of concern to scientists, not parents
4) Is less than the likelihood of a disease epidemic

Which is the primary reason the “no” author believes that all children should receive vaccinations?

1) Our obligation is to protect children, not parents
2) Our obligation is to prevent disease, not side effects
3) Our obligation is to eliminate disease whenever possible
4) Our obligation is to the population, not individuals

Higher Order Questions

APPLY: Which of these situations is most consistent with the “yes” author’s beliefs about a parent’s right to vaccine exemptions?

1) A parent has the right to discipline their child as they see fit
2) A parent has the right to make all decisions for their child
3) A parent has the right to teach religion to their child as they see fit
4) A parent has the right to educate their child as they see fit

ANALYZE: Which author would agree with the following statement? “The ends justify the means.”

1) The “yes” author
2) The “no” author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the “yes” author’s views?

1) Parents are ultimately responsible for their child’s wellbeing
2) Parents always know what is best for their child’s wellbeing
3) The government has no right to interfere with a child’s wellbeing
4) The government has no right to override the wishes of a parent

CREATE: Which education system do you predict the “yes” author would support?

1) A system where the government decides which schools children attend based on ability
2) A system where the government decides which schools children attend based on proximity
3) A system where parents decide which schools children attend based on ability
4) A system where parents decide which schools children attend based on proximity

APPLY: Which of these situations is most consistent with the “no” author’s beliefs about a doctor’s obligation to protect his or her patients?

1) Doctors must do whatever it takes to save a patient’s life, even if treatment is illegal
2) Doctors must do whatever it takes to save a patient’s life, even if treatment is unsafe
3) Doctors must do whatever it takes to save a patient’s life, even if treatment is unethical
4) Doctors must do whatever it takes to save a patient’s life, even if treatment is refused

ANALYZE: Which author would agree with the following statement? “Vaccine development should continue to be a priority of our federal government.”

1) The “yes” author
2) The “no” author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the “no” author’s views?

1) Mandatory vaccination protects the community with only negligible risk
2) Mandatory vaccination protects the community at the risk of children
3) Mandatory vaccination protects the community at the risk of parents
4) Mandatory vaccination protects the community and risk is unavoidable

CREATE: Which opinion regarding a public smoking ban do you predict the “no” author would most likely hold?

1) The ban would benefit smokers, non-smokers, employees, and potential tourists
2) The ban would increase tourism and revenue for restaurants, bars, and casinos
3) The ban would force smokers to stop smoking, thereby improving their health
4) The ban would give families the opportunity to enjoy a smoke-free environment

Should Multiculturalism Be Included In School Curriculum?

Concept Questions

According to the “yes” author, a multicultural education benefits

1) All students, white and minority
2) Teachers and students
3) Teachers, students, and society

4) Minority students

Which is one of the solutions the “yes” author proposes?
1) Identifying how society is biased toward some students
2) Identifying how teachers are biased toward some students
3) Identifying how standardized tests are biased toward some students
4) Identifying how schools are biased toward some students

According to the “yes” author, what is a responsibility of teachers?
1) To build relationships with students’ parents and siblings
2) To understand students’ cultural and linguistic diversity
3) To increase learning by encouraging participation from minority students
4) To treat all white and minority students equally and fairly

Which is the primary reason the “yes” author supports multicultural education?
1) To address the growing diversity of students in our society
2) To encourage teachers to become more sensitive about diversity
3) To develop stronger relationships between white and minority students
4) To overcome social inequalities, such as socioeconomic status

According to the “no” author, what is an outcome of multicultural education?
1) Deep learning about only a few topics
2) Shallow learning about a lot of topics
3) A lack of critical thinking skills
4) The learning of only facts and details

Which is the primary reason the “no” author is against multicultural education?
1) Multicultural education requires the removal of more important topics
2) Multicultural education is too sensitive and emotional for students
3) Multicultural education interferes with the teaching of social studies
4) Multicultural education is a fad that does not enhance student learning

According to the “no” author, why does multiculturalism represent “educational bankruptcy?”
1) It forces teachers to teach more history
2) It hampers standardized test scores
3) It lacks a clear goal or end result
4) It focuses on differences instead of similarities

The “no” author argues that a distinctive American culture is
1) Ubiquitous
2) Detrimental
3) Non-existent
4) Desirable

Higher Order Questions

APPLY: Which of the following programs would the "yes" author most likely support?

1) A program that teaches women how to promote independence and autonomy
2) A program that teaches businesses how to promote community service
3) A program that teaches parents how to promote responsible spending habits
4) A program that teaches college students how to promote social justice

ANALYZE: Which author would agree with the following statement? "Building strong relationships between teachers and students is more important than what is taught."

1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "yes" author's views?

1) Multicultural education is the solution to America's stagnant education system
2) Multicultural education should be required in every public school in the country
3) Multicultural education requires teachers to shift their beliefs and attitudes first
4) Multicultural education is necessary to be successful in today's global economy

CREATE: How do you predict the "yes" author would react to an affirmative action policy at a local college?

1) The "yes" author would support affirmative action because it increases student diversity
2) The "yes" author would not support affirmative action because it promotes unequal treatment of students
3) The "yes" author would support affirmative action because it accounts for past inequalities
4) The "yes" author would not support affirmative action because it emphasizes race instead of academic achievement

APPLY: What type of educational curriculum would the "no" author most likely support?

1) A curriculum that emphasizes fact learning
2) A curriculum that emphasizes in-depth knowledge
3) A curriculum that emphasizes history and literature

4) A curriculum that emphasizes diverse subject matter

ANALYZE: Which author would agree with the following statement? "Education that teaches students knowledge, but not character, morality, and values, is incomplete."

1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "no" author's views?

1) Multicultural education should never be taught in public schools
2) Multicultural education is a waste of classroom time and resources
3) Multicultural education harms students over the long-term
4) Multicultural education is less effective than a traditional curriculum

CREATE: How do you predict the "no" author would react to a multicultural training program for employees at a local business?

1) The "no" author would not support the program because multicultural training in the workplace is not valuable
2) The "no" author would support the program because a multicultural workplace atmosphere would increase profits
3) The "no" author would support the program because it is for adults, not students, so it wouldn't negatively effect learning
4) The "no" author would not support the program because it is not the employer's responsibility to encourage multiculturalism

Will Biotech Solve Africa's Food Problems?

Concept Questions

Which is one of the solutions the "yes" author proposes?

1) Reallocate funds from anti-biotech activists toward more research
2) Drastically improve and increase the number of roads to food markets
3) Conduct more biotech research by diverse experts around the world
4) Increase the production of genetically modified crops and fertilizers

What type of farming method does the "yes" author support?

1) An academic model with researcher-to-farmer diffusion of information
2) A diversity model where farmers use biotech, pesticides, and fertilizers
3) A cooperative model composed of farmers, livestock owners, and researchers
4) A self-sufficient model where farmers conduct research and produce food

According to the “yes” author, what is one downside of foreign aid for African farmers?

1) Foreign aid is insufficient to adequately sustain African farmers
2) Foreign aid is expensive and cannot continue for much longer
3) African farmers would become dependent on foreign aid
4) Foreign aid offers short-term, but not permanent, solutions

Which is the primary reason the “yes” author supports biotech in Africa?

1) To increase food production more effectively using research
2) To increase food production in Africa, by Africans themselves
3) To increase food production and improve Africa’s economy
4) To increase food production and reduce worldwide hunger

What is one benefit of the fallow period technique the “no” author supports?

1) It reduces pesticide use
2) It is cheap and accessible
3) It helps local economies
4) It is safer and healthier

According to the “no” author, how might an increase in food production worsen Africa’s food problems?

1) It could increase profits for non-Africans only
2) It could increase the use of pesticides and other chemicals
3) It could increase supply without increasing demand
4) It could decrease the amount of land available for farming

Which is the primary reason the “no” author is against biotech in Africa?

1) Biotech benefits corporations, not farmers
2) Biotech has failed to work in the past
3) Biotech harms the environment
4) Biotech has failed to develop long-term solutions

According to the “no” author, why are Africa’s farmers the “hungriest occupation on Earth?”

1) Farmers are unable to produce enough food
2) Farmers have trouble selling the food they produce
3) Farmers fail to use technology correctly
4) Farmers don’t make enough profit when selling food

Higher Order Questions

APPLY: What is a potential benefit of the farming method the “yes” author supports?

1) More biotech research could be conducted in a limited amount of time
2) Male and female farmers would have an equal role in food production

3) Farmers could decrease their use of pesticides and increase profits
4) Inclusion of different age groups may enhance the diffusion of knowledge

ANALYZE: Which author would agree with the following statement? "African farmers need to be able to produce food on their own, without foreign help."
1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "yes" author's views?
1) Funds spent on farming techniques other than biotech is a waste of money
2) Information sharing between farmers and researchers, and vice versa, is critical
3) Food production is vital, and we must do everything we can to increase yield
4) Farmers need to increase their profits, and using biotech research is the solution

CREATE: Which effort do you predict the "yes" author would support if there were a pest infestation in Africa?
1) A grant for research to be conducted by African farmers and scientists
2) A change in pesticide type, based on research from another country
3) A class for farmers to learn about pest control techniques from scientists
4) A supply of research-based fertilizer developed in another part of Africa

APPLY: What would happen if African farmers exported their food to other continents?
1) Farmers would be able to do this individually, save money, and make more profit
2) Farmers would still have the same issues of transportation, cost, and demand
3) Farmers could profit without having to use biotech to increase food production
4) Farmers would have a negative effect on the local African economy

ANALYZE: Which author would agree with the following statement? "Hunger is getting worse in Africa because of a lack of resources and money."
1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the “no” author’s views?

1) Selling directly to African consumers is the solution to Africa’s food problems
2) Biotechnology is expensive, under developed, and bad for the environment
3) African farmers should never use chemicals, which only benefit corporations
4) Empowering farmers through the use of high-yield techniques is Africa’s solution

CREATE: How do you predict the “no” author would react to the organic food movement?

1) Good, because organic food is easy to grow without using chemicals
2) Great, because organic food is natural and healthy
3) Just okay, because organic food is costly to both the consumer and farmer
4) Not good, because organic food is not always produced locally

Should We Continue to Study Sex Differences?

Concept Questions

According to the “yes” author, why might some people be anxious about sex differences between men and women?

1) Because they imply that women should be treated different from men
2) Because they imply that women have always been treated unequally
3) Because they imply that women will always be different from men
4) Because they imply that women have always been inferior to men

According to the “yes” author, feminists are against sex difference research because

1) Feminists strive to highlight gender similarities instead of gender differences
2) Feminists are fundamentally against the practice of comparing men to women
3) Feminists only support research that shows that women are better than men
4) Feminists argue that sex difference research oppresses and offends women

The “yes” author argues that the current stereotype about women is

1) Less positive than the stereotype for men
2) Almost the same as the stereotype for men
3) More positive than the stereotype for men

4) More negative than the stereotype for men

Which is the primary reason the “yes” author believes that we should continue to study sex differences?

1) Because this area of research is ripe for exciting discoveries and theory testing
2) Because women can address inequalities and strive to achieve equal treatment
3) Because we haven’t conducted enough research yet to draw any conclusions
4) Because we will be better informed and can adjust our unequal political agenda

According to the “no” author, a gender difference ideology

1) Harms women and only benefits men
2) Causes an increase in gender differences
3) Values one gender over the other
4) Focuses on the conditions of inequality

According to the “no” author, what is one way to combat a gender difference ideology?

1) Eliminate funding for sex difference research altogether
2) Gather evidence of similarities between men and women
3) Celebrate the differences between men and women
4) Educate the public about minority genders and races

The “no” author argues that a gender difference ideology is a source of

1) Confusion and stress
2) Hatred and oppression
3) Political indifference
4) Inaccurate stereotypes

Which is the primary reason the “no” author believes that we should stop studying sex differences?

1) This research only serves a philosophical purpose
2) This research only serves a scientific purpose
3) This research only serves a cultural purpose
4) This research only serves a political purpose

Higher Order Questions

APPLY: Which of these situations is most consistent with the “yes” author’s beliefs about the purpose of studying sex differences?

1) A study finds that men are better than women at math, so a professor gives women a few extra points on a math test

2) A study finds that men are better at engineering than women, so women majoring in engineering work harder
3) A study finds that women are better than men at management, so companies hire more women for leadership positions
4) A study finds that women are better than men at saving money, so the government gives tax incentives to men

ANALYZE: Which author would agree with the following statement? "Study of biological sex differences may portray women as inferior."
1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "yes" author's views?
1) Research on sex differences is a small, but growing, field of study
2) Research on sex differences yields no negative effects for women
3) Research on sex differences is too important and valuable to abandon
4) Research on sex differences has a unwarranted negative reputation

CREATE: How do you predict the "yes" author would react to a utopian society in which men and women were treated the same?
1) The "yes" author would be supportive, but would still encourage sex difference research
2) The "yes" author would be surprised, because men and women can't be treated the same
3) The "yes" author would be disappointed, because sex difference research would be ignored
4) The "yes" author would be excited, and would no longer conduct sex difference research

APPLY: Which of these child-rearing techniques would the "no" author most likely support?
1) Treating boys and girls completely equal in every way possible
2) Fostering an understanding of both similarities and differences
3) Fostering an understanding of how boys and girls are different
4) Fostering an understanding of how boys and girls are similar

ANALYZE: Which author would agree with the following statement? "Sex differences are a result of nature or genes, not nurture or environment."
1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "no" author's views?

1) A gender difference ideology is destructive and must be eliminated from our society
2) A gender difference ideology only serves to improve the position of men, not women
3) A gender difference ideology ignores the influence of culture, context, and history on women
4) A gender difference ideology is a social construction, promoted by male scientists

CREATE: How do you predict the "no" author would respond if the U.S. government funded an increase in sex difference research?

1) The "no" author would argue that the government is providing a service to men
2) The "no" author would argue that the government is providing a disservice to society
3) The "no" author would argue that the government is providing a disservice to children
4) The "no" author would argue that the government is providing a service to politicians

Was Abraham Lincoln America's Greatest President?

Concept Questions

According to the "yes" author, why might Lincoln have been unfit to be President?

1) He lacked administrative experience
2) He was too arrogant and demanding
3) He lacked experience as a war general
4) He was too controversial and unpopular

How did Lincoln manage to pass the emancipation amendment?

1) He insisted upon equal rights under the Declaration of Independence
2) He included some restrictions on slaves, which benefited the North
3) He promoted the amendment as the only way to end the Civil War
4) He persisted until members of Congress agreed it was the right thing to do

According to the "yes" author, why did Lincoln support emancipation?

1) He wanted to guarantee equal rights to all people
2) He wanted to remain consistent with his moral convictions
3) He wanted to guarantee the right of self-government
4) He wanted to demonstrate the power of his influence

What was one negotiation that Lincoln approved?

1) He agreed to support a bill he previously opposed
2) He made exceptions for certain slave owners
3) He provided jobs for relatives of congressmen
4) He provided additional money to Republicans

Lincoln emancipated slaves,

1) Although the situation required compromise
2) And he was solely responsible for this outcome
3) Which he always expected to accomplish
4) With overwhelming support from Congress

According to the "no" author, why did Lincoln institute heavy taxes?

1) To help the U.S. economy during wartime
2) To provide subsidies for big corporations
3) To pay for military resources and weapons
4) To distribute income from the rich to the poor

Which is the primary reason the "yes" author supported Lincoln?

1) Lincoln overcame adversity and depression
2) Lincoln was an effective general and ended the war
3) Lincoln was a warrior for the American dream
4) Lincoln passed the emancipation amendment

Which is the primary reason the "no" author was against Lincoln?

1) Lincoln was dishonest and corrupt
2) Lincoln supported big corporations
3) Lincoln was a power hungry dictator
4) Lincoln touted his agenda as morally superior

Higher Order Questions

APPLY: If Lincoln tried to pass a bill, but he did not receive enough votes in Congress, what would he mostly likely have done next?

1) He would have appealed directly to all voters
2) He would have appealed to members of Congress
3) He would have appealed to only those affected by the bill
4) He would have revised and resent the bill

ANALYZE: Which author would agree with the following statement? "Using one's power of persuasion, even if you have to bend the rules, is sometimes necessary."

1) The "yes" author
2) The "no" author
3) Both authors

4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "yes" author's views?

1) Considering Lincoln's depression, his accomplishments are impressive
2) Lincoln could persuade others, though he was unwilling to be persuaded
3) Lincoln often followed his convictions and he accomplished a great deal
4) Considering Lincoln's beliefs about emancipation, he conquered an uphill battle

CREATE: Which of the following policy initiatives do you predict Lincoln would most likely encourage if he were alive today?

1) Bailouts for car companies and banks, because he supported corporations in the past
2) Equal rights for homosexuals, because he supported equal rights for slaves
3) An end to overseas wars, because he had an aversion to bloodshed and violence
4) Limitation of government mandates, because he supported the right to self-government

APPLY: Considering Lincoln's preference regarding federal powers during the Civil War, Lincoln most likely would have

1) Supported George W. Bush's use of federal powers following September 11th, 2001
2) Disagreed with George W. Bush's use of federal powers following September 11th, 2001
3) Encouraged George W. Bush to seek Congressional approval for the use of federal powers following September 11th, 2001
4) Encouraged George W. Bush to negotiate with international leaders following September 11th, 2001

ANALYZE: Which author would agree with the following statement? "Lincoln's true beliefs were not always in accordance with the outcome of a situation."

1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "no" author's views?

1) Even though Lincoln did what was best only for the North, he still deserves praise for emancipating slaves
2) Lincoln was a cheating, greedy President, but he was responsible for some positive results

3) Lincoln was responsible for many poor decisions, and America was worse off because of Lincoln
4) Lincoln could not be trusted, as he deceived Americans in every situation and at every turn

CREATE: Which of the following governance strategies do you predict Lincoln would most likely support if he were alive today?
1) The obligation to always do what is morally right
2) The necessary use of compromise in politics
3) The responsibility to follow the Constitution literally
4) The commitment to protect the American people

Is the Superfund Program Successfully Protecting the Environment from Hazardous Wastes?

Concept Questions

Before 1995, how did the Superfund program assess future health hazards?
1) They assumed future residential use at the site
2) They assumed future commercial use at the site
3) They used projected exposure patterns
4) They used a risk-based hierarchy system

What is an example of a remedy solution for a low-risk site?
1) Removal of contaminated soil
2) Incineration of waste and toxins
3) Prohibition of drinking water wells
4) Treatment of contaminated water

According to the “yes” author, what are two areas in which the EPA has made progress?
1) Scientific research and treatment strategies
2) Scientific research and cleanup effectiveness
3) Risk assessments and cleanup effectiveness
4) Risk assessments and treatment strategies

Early Superfund cleanup efforts were focused on which strategy?
1) Source control treatment
2) Excavation and disposal
3) Soil vapor extraction
4) Acid water containment

According to the “no” author, obstacles for the Superfund program include
1) Political opposition to cleanup efforts

2) Conflicts with other environmental agencies
3) Resident opposition to cleanup efforts
4) Lack of technology and research

Why is contamination at Tar Creek spreading?
1) EPA failed to improve the air quality
2) EPA failed to invest money in the cleanup effort
3) EPA failed to contain gallons of acid water
4) EPA failed to truck out toxic dirt

How did Tar Creek become contaminated?
1) An abundance of toxic waste dumping
2) An abundance of mine shaft drilling
3) An abundance of noxious chemical use
4) An abundance of harmful radiation use

Consideration of implementation risks is important, because otherwise
1) Contamination will increase and spread
2) Cleanup will become impractical
3) Superfund sites will be ignored
4) Some techniques will be too costly

Higher Order Questions

APPLY: According to the risk-based hierarchy system, for a cancer risk of 1 in 1,000,000 at a Superfund site, which remedy solution is most appropriate?
1) Removal of waste
2) Treatment of waste
3) Prohibiting land/water use
4) Restricting land/water use

ANALYZE: Which author would agree with the following statement? "An increase in federal funding for Superfund would be worthwhile, because cleanup of waste sites is crucial."
1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "yes" author's views?
1) The Superfund has made great progress recently
2) The Superfund program has always worked really well
3) The Superfund program needs a lot of improvement
4) The Superfund program is both effective and ineffective

CREATE: If the "yes" author had $100 million to donate to the Superfund program, what do you predict he or she would encourage?

1) An increase in waste removal and disposal efforts
2) Greater development of flexible treatment options
3) Greater development of incineration techniques
4) An increase in research conducted at Superfund sites

APPLY: If the EPA trucked out the pollution from all of the homes and schools in Tar Creek, what would be the result?

1) The "no" author would be satisfied with the newly cleaned Tar Creek
2) The other waste problems in Tar Creek would be fixed
3) There would still be long-term health consequences from the waste
4) Tar Creek residents' would no longer require kidney dialysis

ANALYZE: Which author would agree with the following statement: "EPA buyouts and settlements are not solutions, even if they increase the amount of money in the Superfund program."

1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "no" author's views?

1) The Superfund has made great progress recently
2) The Superfund program needs a lot of improvement
3) The Superfund program is both effective and ineffective
4) The Superfund program has always been ineffective

CREATE: If Superfund received an influx of funding, how do you predict the "no" author would react?

1) Good, but Superfund may still be ineffective because of partisan politics
2) Great, budget problems are the main reason for Superfund's ineffectiveness
3) Useless, money won't make Superfund more effective
4) Fine, but most of the money should go toward cleaning up Tar Creek

Did World War II Liberate American Women?

Concept Questions

According to the "yes" author, what laid the groundwork for the women's liberation movement?

1) New expectations from husbands and other men

2) Insistence from feminists that women change
3) Magazine articles promoting a transformation
4) Desire of women to test and prove themselves

According to the "yes" author, how did earning money contribute to the women's liberation movement?

1) It gave women the means to make spending decisions
2) It gave women the confidence to demand greater authority
3) It gave women the independence to challenge their husbands
4) It gave women the opportunity to leave the household

The "yes" author argues that the lag between evolving ideas regarding women and the actual practice of these ideas is

1) Typical
2) Unacceptable
3) Unexpected
4) Frustrating

Which is the primary reason the "yes" author believes that women's wartime experience played a vital role in a liberation movement?

1) Because of the wartime need for women to serve as heads of households
2) Because of the wartime need for women to earn money for their family
3) Because of the wartime need for women to undertake employment duties
4) Because of the wartime need for women to fulfill both roles of mother and father

According to the "no" author, why were few women able to keep their jobs after the war?

1) Women failed to keep up the same workplace efficiency as men were able to
2) Women were encouraged to give their jobs back to men when they returned
3) Women were asked to focus on their family, instead of work, when men returned
4) Women failed to receive any encouragement from men to continue working

According to the "no" author, why were single women encouraged to find husbands after the war?

1) Because society wanted to encourage a "baby boom"
2) Because society sought to keep them out of the workforce
3) Because society did not want them to enter the workforce
4) Because society had doubts about single women's motives

The “no” author argues that wartime primarily instilled

1) Ambivalence in women
2) Dependency in women
3) Patriotism in women
4) Responsibility in women

Which is the primary reason the “no” author disagrees that women’s wartime experience contributed to a liberation movement?

1) Popular culture encouraged women to retain their femininity
2) Popular culture recruited women into the labor force temporarily
3) Popular culture continued to reinforce traditional gender roles
4) Popular culture returned its focus to men when they came home

Higher Order Questions

APPLY: How might the “yes” author react to women who choose to care for children full-time vs. women who choose to work full-time?

1) The “yes” author would support women who work full-time because they are financially independent
2) The “yes” author would support women who care for children full-time because raising a family is an important role
3) The “yes” author would support both types of women for asserting their preference and choosing their own lifestyle
4) The “yes” author would support women who split their time between the two and achieve a family-work balance

ANALYZE: Which author would agree with the following statement? “World War II restructured society and paved the way for transformation of traditional gender roles.”

1) The “yes” author
2) The “no” author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the “yes” author’s views?

1) Gender roles have changed because mothers have influenced their daughters
2) Gender roles have changed because individuals pushed for social transformation
3) Gender roles have changed because of an increasing societal value of equality
4) Gender roles have changed because of an accumulation of similar experiences

CREATE: Currently, 53% of the workforce is male, and 47% of the workforce is female. How do you predict the "yes" author would respond to this current inequality?

1) Progress is too slow, hiring practices and incentives for women must improve
2) Progress takes time, but at least the situation is better than it was after WWII
3) Progress is better than expected, demonstrating the large influence of WWI
4) Progress has reached its peak, the current situation is as good as possible

APPLY: How might the "no" author describe the social climate for men when they returned from the war?

1) Men returned to the same social climate at home, but not at work
2) Men returned to the same social climate at both home and work
3) Men returned to different social climates at both home and work
4) Men returned to a different social climate at home, but not at work

ANALYZE: Which author would agree with the following statement? "Women's wartime experience had a large influence on their children."

1) The "yes" author
2) The "no" author
3) Both authors
4) Neither author

EVALUATE: Which statement is an accurate evaluation or summary of the "no" author's views?

1) Although women were eager to undertake new roles, society was not prepared to allow women to do so, at least not permanently
2) Although women helped maintain the wartime economy by working, women appropriately returned to their household duties
3) Although women were recruited to join the workforce, politicians urged them to maintain their household duties simultaneously
4) Although women had the opportunity to break out of traditional gender roles, they failed to take advantage of the situation

CREATE: How do you predict the "no" author would react to a present-day military draft of men?

1) Women's wartime roles would still be viewed as gender-specific because society today is the same as it was in the past
2) Women's wartime roles would still be viewed as gender-specific because wartime always limits the options of women
3) Women's wartime roles would be viewed as gender-neutral because society is more gender-neutral today than in the past
4) Women's wartime roles would be viewed as gender-neutral because wartime does not present the same economic difficulties

Appendix D

Rephrased Concept and Higher Order Questions Used in Experiments 1 and 2, Session 2

Note. Correct answers are underlined, and the type of higher order question (apply, analyze, evaluate, or create) is indicated below for illustrative purposes, but correct answers and higher order types were not revealed to subjects during testing.

Does Welfare Do More Harm Than Good?

Rephrased Concept Questions

The "yes" author proposes that the government
1) Keep all welfare programs, but reduce spending within each program
2) Eliminate some welfare programs and increase spending for remaining programs
3) Eliminate all welfare programs in the United States
4) Eliminate Medicaid, but keep all veteran benefits

According to the "yes" author, welfare programs are expensive because
1) There are too many recipients and not enough taxpayers
2) Recipients are dependent and require a lot of assistance
3) A great deal of staff are needed to administer the programs
4) There is little connection between taxpayers and legislators

According to the "yes" author, the Social Security program is problematic because
1) Taxes raised are not enough to keep the program sustainable
2) Taxes should not be required because older adults can take care of themselves
3) Taxes raised are not enough to help the federal government
4) Taxes should not be required from younger adults to pay for older adults

The "yes" author is against welfare programs, largely because
1) Welfare programs are too expensive for taxpayers
2) Welfare programs don't benefit recipients or taxpayers
3) Welfare programs are not the government's responsibility
4) Welfare programs create dependence for recipients

The "no" author supports welfare programs because
1) They are affordable and feasible
2) They help everyone, not just recipients

3) They eradicate discrimination
4) They help support local communities

According to the "no" author, taxation is required in order to
1) Provide citizens with services they can't pay for on their own
2) Provide the government with means to improve society
3) Provide the government a way to act on citizens' behalf
4) Provide citizens a way to support their government

The "no" author supports welfare programs, largely because
1) They improve, not hinder, economic growth
2) They are a good investment of taxpayer money
3) They create independence, not dependence
4) They are the government's responsibility

The "no" author believes that a free market system
1) Is insufficient to provide equality for citizens
2) Is the only alternative to welfare programs
3) Helps make welfare programs even stronger
4) Can address problems of discrimination

Rephrased Higher Order Questions

APPLY: If there were no welfare programs in the future, the "yes" author would expect
1) A society in which no one would be required to pay taxes
2) A society in which all individuals are self-reliant and independent
3) A society in which all individuals are treated equally
4) A society in which there would be no role for the government

ANALYZE: Which author would agree with the following statement? "A government that helps society is admirable."
1) The "no" author
2) Neither author
3) The "yes" author
4) Both authors

EVALUATE: The "yes" author would agree with which statement?
1) Welfare programs could work, but they rarely meet the needs of the people
2) Welfare programs waste taxpayer money on people who don't really need help
3) Welfare programs can never work, because they are always too expensive

4) Welfare programs are harmful, because they make bad situations even worse

CREATE: If the "yes" author became unemployed and needed welfare assistance,

1) The "yes" author might accept government assistance, but would try to find a new job first
2) The "yes" author would not accept government assistance, but would seek help from local organizations
3) The "yes" author would not accept government assistance, but would try to find a new job
4) The "yes" author might accept government assistance, but would seek help from local organizations first

APPLY: The "no" author would support which of the following governance strategies?

1) Governments around the world are obligated to help poor countries
2) Governments around the world are obligated to help countries that reciprocate
3) Governments around the world are obligated to help when asked
4) Governments around the world are obligated to help all countries

ANALYZE: Which author would agree with the following statement? "To spur economic growth, governments should invest in people."

1) Both authors
2) Neither author
3) The "no" author
4) The "yes" author

EVALUATE: The "no" author would agree with which statement?

1) The government's primary role is advancing liberty
2) The government's primary role is advancing equality
3) The government's primary role is advancing morality
4) The government's primary role is advancing security

CREATE: The "no" author would support which tax and spending structure?

1) Higher taxes for the rich, lower taxes for the poor; less spending on the rich, more spending on the poor
2) Equal taxation of all Americans; less spending on the rich, more spending on the poor
3) Higher taxes for the rich, lower taxes for the poor; more spending on the rich, less spending on the poor
4) Equal taxation of all Americans; equal spending on all Americans

Should Parents Be Allowed to Opt Out of Vaccinating Their Children?

Rephrased Concept Questions

According to the "yes" author, vaccination should be optional because
1) Vaccines do more harm than good
2) We can't screen out vulnerable children
3) Vaccines are prohibited by most religions
4) Research has verified its ineffectiveness

According to the "yes" author, all parents have
1) The right to decide on behalf of their children
2) An obligation to decide on behalf of their children
3) A responsibility to decide on behalf of their children
4) The option to decide on behalf of their children

The "yes" author argues that more vaccine research should be conducted because we need to
1) Determine the long-term effect of vaccines
2) Determine the effectiveness of vaccines
3) Determine the mechanism behind vaccines
4) Determine the side effects from vaccines

The "yes" author believes that parents should be able to opt out of vaccination, largely because
1) Vaccination practices lack solid research
2) Vaccination for all children is too simplistic
3) Vaccination has the potential to cause death
4) Vaccination costs outweigh the benefits

According to the "no" author, autism diagnoses have increased, not because of the measles vaccine, but because of
1) A lack of understanding of autism
2) Poor childhood nutrition or immunity
3) Some being more at-risk than others
4) A change in the diagnostic definition

According to the "no" author, if parents were allowed to opt out of vaccination,
1) This would be malpractice, and against state and federal law
2) This would be a decision for medical professionals, not politicians
3) This would increase, not decrease, danger to the population
4) This would only lead to more and more parents opting out

The "no" author argues that vaccine risk
1) Should be of concern to scientists, not parents
2) Is a possibility with any medical procedure

3) Is less than the likelihood of a disease epidemic
4) Is too small to be of concern to the community

The “no” author believes that all children should receive vaccinations, largely because

1) Our obligation is to the population, not individuals
2) Our obligation is to prevent disease, not side effects
3) Our obligation is to protect children, not parents
4) Our obligation is to eliminate disease whenever possible

Rephrased Higher Order Questions

APPLY: The “yes” author’s beliefs about a parent’s right to vaccine exemptions are most consistent with which of these situations?

1) A parent has the right to teach religion to their child as they see fit
2) A parent has the right to educate their child as they see fit
3) A parent has the right to discipline their child as they see fit
4) A parent has the right to make all decisions for their child

ANALYZE: Which author would agree with the following statement? “The achieved outcome is more important than the process along the way.”

1) The “no” author
2) Neither author
3) Both authors
4) The “yes” author

EVALUATE: The “yes” author would agree with which statement?

1) Parents always know what is best for their child’s wellbeing
2) The government has no right to override the wishes of a parent
3) Parents are ultimately responsible for their child’s wellbeing
4) The government has no right to interfere with a child’s wellbeing

CREATE: The “yes” author would support which of these education systems?

1) A system where the government decides which schools children attend based on proximity
2) A system where parents decide which schools children attend based on proximity
3) A system where the government decides which schools children attend based on ability
4) A system where parents decide which schools children attend based on ability

APPLY: The “no” author’s beliefs about a doctor’s obligation to protect his or her patients are most consistent with which of these situations?

1) Doctors must do whatever it takes to save a patient’s life, even if treatment is unsafe
2) Doctors must do whatever it takes to save a patient’s life, even if treatment is refused
3) Doctors must do whatever it takes to save a patient’s life, even if treatment is illegal
4) Doctors must do whatever it takes to save a patient’s life, even if treatment is unethical

ANALYZE: Which author would agree with the following statement? “Our federal government should continue to prioritize vaccine development.”

1) The “yes” author
2) Both authors
3) Neither author
4) The “no” author

EVALUATE: The “no” author would agree with which statement?

1) Mandatory vaccination protects the community at the risk of parents
2) Mandatory vaccination protects the community with only negligible risk
3) Mandatory vaccination protects the community and risk is unavoidable
4) Mandatory vaccination protects the community at the risk of children

CREATE: Regarding a public smoking ban, the “no” author would argue that

1) The ban would give families the opportunity to enjoy a smoke-free environment
2) The ban would force smokers to stop smoking, thereby improving their health
3) The ban would increase tourism and revenue for restaurants, bars, and casinos
4) The ban would benefit smokers, non-smokers, employees, and potential tourists

Should Multiculturalism Be Included In School Curriculum?

Rephrased Concept Questions

According to the “yes” author, a multicultural education helps which group of people?

1) Minority students
2) Teachers, students, and society
3) Teachers and students
4) All students, white and minority

The “yes” author argues that a multicultural education should include

1) Identifying how schools are biased toward some students
2) Identifying how society is biased toward some students
3) Identifying how standardized tests are biased toward some students
4) Identifying how teachers are biased toward some students

According to the “yes” author, teachers are responsible for

1) Treating all white and minority students equally and fairly
2) Increasing learning by encouraging participation from minority students
3) Building relationships with students’ parents and siblings
4) Understanding students’ cultural and linguistic diversity

The “yes” author supports multicultural education, largely because it

1) Develops stronger relationships between white and minority students
2) Overcomes social inequalities, such as socioeconomic status
3) Addresses the growing diversity of students in our society
4) Encourages teachers to become more sensitive about diversity

According to the “no” author, a multicultural education typically results in

1) A lack of critical thinking skills
2) Deep learning about only a few topics
3) Shallow learning about a lot of topics
4) The learning of only facts and details

The “no” author is against multicultural education, largely because

1) Multicultural education is a fad that does not enhance student learning
2) Multicultural education is too sensitive and emotional for students
3) Multicultural education interferes with the teaching of social studies
4) Multicultural education requires the removal of more important topics

According to the “no” author, multiculturalism represents an “educational bankruptcy” because

1) It lacks a clear goal or end result
2) It hampers standardized test scores
3) It forces teachers to teach more history
4) It focuses on differences instead of similarities

The “no” author argues that promotion of a unique American culture is

1) Detrimental
2) Desirable
3) Non-existent
4) Ubiquitous

Rephrased Higher Order Questions

APPLY: The “yes” author would most likely support which of the following programs?

1) A program that teaches parents how to promote responsible spending habits
2) A program that teaches women how to promote independence and autonomy
3) A program that teaches college students how to promote social justice
4) A program that teaches businesses how to promote community service

ANALYZE: Which author would agree with the following statement? “Content taught in classrooms is less important than building strong relationships between teachers and students.”

1) Both authors
2) The “no” author
3) Neither author
4) The “yes” author

EVALUATE: The “yes” author would agree with which statement?

1) Multicultural education should be required in every public school in the country
2) Multicultural education requires teachers to shift their beliefs and attitudes first
3) Multicultural education is necessary to be successful in today’s global economy
4) Multicultural education is the solution to America’s stagnant education system

CREATE: If an affirmative action policy were implemented at a local college,

1) The “yes” author would not support affirmative action because it promotes unequal treatment of students
2) The “yes” author would support affirmative action because it accounts for past inequalities
3) The “yes” author would support affirmative action because it increases student diversity
4) The “yes” author would not support affirmative action because it emphasizes race instead of academic achievement

APPLY: The “no” author would most likely support which of these curricula?

1) A curriculum that emphasizes diverse subject matter
2) A curriculum that emphasizes fact learning
3) A curriculum that emphasizes history and literature
4) A curriculum that emphasizes in-depth knowledge

ANALYZE: Which author would agree with the following statement? "Education that teaches students content, but not disposition, ethics, and ideals, is incomplete."

1) Both authors
2) The "no" author
3) The "yes" author
4) Neither author

EVALUATE: The "no" author would agree with which statement?

1) Multicultural education is less effective than a traditional curriculum
2) Multicultural education should never be taught in public schools
3) Multicultural education harms students over the long-term
4) Multicultural education is a waste of classroom time and resources

CREATE: If a multicultural training program for employees were implemented at a local business,

1) The "no" author would support the program because a multicultural workplace atmosphere would increase profits
2) The "no" author would support the program because it is for adults, not students, so it wouldn't negatively effect learning
3) The "no" author would not support the program because it is not the employer's responsibility to encourage multiculturalism
4) The "no" author would not support the program because multicultural training in the workplace is not valuable

Will Biotech Solve Africa's Food Problems?

Rephrased Concept Questions

The "yes" author proposed which potential solution to Africa's food problems?

1) Conduct more biotech research by diverse experts around the world
2) Increase the production of genetically modified crops and fertilizers
3) Reallocate funds from anti-biotech activists toward more research
4) Drastically improve and increase the number of roads to food markets

The "yes" author supports which farming method?

1) A self-sufficient model where farmers conduct research and produce food
2) A diversity model where farmers use biotech, pesticides, and fertilizers
3) An academic model with researcher-to-farmer diffusion of information
4) A cooperative model composed of farmers, livestock owners, and researchers

According to the “yes” author, foreign aid for African farmers presents which potential obstacle?

1) African farmers would become dependent on foreign aid
2) Foreign aid offers short-term, but not permanent, solutions
3) Foreign aid is expensive and cannot continue for much longer
4) Foreign aid is insufficient to adequately sustain African farmers

The “yes” author supports biotech in Africa, largely because it may

1) Increase food production in Africa, by Africans themselves
2) Increase food production and reduce worldwide hunger
3) Increase food production more effectively using research
4) Increase food production and improve Africa’s economy

The “no” author supports a fallow period technique, because

1) It is safer and healthier
2) It helps local economies
3) It reduces pesticide use
4) It is cheap and accessible

According to the “no” author, an increase in food production might hurt, not help, Africa’s food problems because

1) It could increase supply without increasing demand
2) It could increase the use of pesticides and other chemicals
3) It could decrease the amount of land available for farming
4) It could increase profits for non-Africans only

The “no” author is against biotech in Africa, largely because

1) Biotech harms the environment
2) Biotech benefits corporations, not farmers
3) Biotech has failed to develop long-term solutions
4) Biotech has failed to work in the past

According to the “no” author, Africa’s farmers are the “hungriest occupation on Earth” because

1) Farmers have trouble selling the food they produce
2) Farmers fail to use technology correctly
3) Farmers don’t make enough profit when selling food
4) Farmers are unable to produce enough food

Rephrased Higher Order Questions

APPLY: The farming method the “yes” author supports may be helpful because

1) Farmers could decrease their use of pesticides and increase profits
2) More biotech research could be conducted in a limited amount of time
3) Inclusion of different age groups may enhance the diffusion of knowledge

4) Male and female farmers would have an equal role in food production

ANALYZE: Which author would agree with the following statement? "Instead of relying on foreign help, African farmers must produce food independently."

1) Neither author
2) Both authors
3) The "no" author
4) The "yes" author

EVALUATE: The "yes" author would agree with which statement?

1) Information sharing between farmers and researchers, and vice versa, is critical
2) Food production is vital, and we must do everything we can to increase yield
3) Farmers need to increase their profits, and using biotech research is the solution
4) Funds spent on farming techniques other than biotech is a waste of money

CREATE: If there were a pest infestation in Africa, the "yes" author would encourage

1) A supply of research-based fertilizer developed in another part of Africa
2) A class for farmers to learn about pest control techniques from scientists
3) A change in pesticide type, based on research from another country
4) A grant for research to be conducted by African farmers and scientists

APPLY: If African farmers exported their food to other continents,

1) Farmers would still have the same issues of transportation, cost, and demand
2) Farmers would have a negative effect on the local African economy
3) Farmers could profit without having to use biotech to increase food production
4) Farmers would be able to do this individually, save money, and make more profit

ANALYZE: Which author would agree with the following statement? "A lack of resources and money are making Africa's food problems worse."

1) The "no" author
2) The "yes" author
3) Neither author
4) Both authors

EVALUATE: The "no" author would agree with which statement?

1) Empowering farmers through the use of high-yield techniques is Africa's solution

2) Biotechnology is expensive, under developed, and bad for the environment
3) African farmers should never use chemicals, which only benefit corporations
4) Selling directly to African consumers is the solution to Africa's food problems

CREATE: How might the "no" author feel regarding the current organic food movement?

1) Great, because organic food is natural and healthy
2) Just okay, because organic food is costly to both the consumer and farmer
3) Not good, because organic food is not always produced locally
4) Good, because organic food is easy to grow without using chemicals

Should We Continue to Study Sex Differences?

Rephrased Concept Questions

According to the "yes" author, some people may be anxious about sex differences between men and women because

1) They imply that women have always been treated unequally
2) They imply that women have always been inferior to men
3) They imply that women will always be different from men
4) They imply that women should be treated different from men

According to the "yes" author, feminists are opposed to sex difference research because

1) Feminists argue that sex difference research oppresses and offends women
2) Feminists are fundamentally against the practice of comparing men to women
3) Feminists only support research that shows that women are better than men
4) Feminists strive to highlight gender similarities instead of gender differences

The "yes" author argues that the current stereotype about men is

1) More negative than the stereotype for women
2) Almost the same as the stereotype for women
3) More positive than the stereotype for women
4) Less positive than the stereotype for women

The “yes” author believes that we should continue to study sex differences, largely because

1) We haven’t conducted enough research yet to draw any conclusions
2) We will be better informed and can adjust our unequal political agenda
3) Women can address inequalities and strive to achieve equal treatment
4) This area of research is ripe for exciting discoveries and theory testing

According to the “no” author, research on gender differences

1) Values one gender over the other
2) Focuses on the conditions of inequality
3) Causes an increase in gender differences
4) Harms women and only benefits men

According to the “no” author, gender difference research findings can be challenged by

1) Educating the public about minority genders and races
2) Eliminating funding for sex difference research altogether
3) Celebrating the differences between men and women
4) Gathering evidence of similarities between men and women

The “no” author argues that gender difference research is responsible for

1) Inaccurate stereotypes
2) Confusion and stress
3) Political indifference
4) Hatred and oppression

The “no” author believes that we should stop studying sex differences, largely because

1) This research only serves a scientific purpose
2) This research only serves a philosophical purpose
3) This research only serves a political purpose
4) This research only serves a cultural purpose

Rephrased Higher Order Questions

APPLY: The “yes” author’s beliefs about the purpose of studying sex differences are most consistent with which of these situations?

1) A study finds that women are better than men at management, so companies hire more women for leadership positions
2) A study finds that women are better than men at saving money, so the government gives tax incentives to men
3) A study finds that men are better than women at math, so a professor gives women a few extra points on a math test
4) A study finds that men are better at engineering than women, so women majoring in engineering work harder

ANALYZE: Which author would agree with the following statement? "Women may be depicted as inferior when studying biological sex differences."

1) Neither author
2) Both authors
3) The "no" author
4) The "yes" author

EVALUATE: The "yes" author would agree with which statement?

1) Research on sex differences has an unwarranted negative reputation
2) Research on sex differences is too important and valuable to abandon
3) Research on sex differences is a small, but growing, field of study
4) Research on sex differences yields no negative effects for women

CREATE: Regarding a utopian society in which men and women are treated the same,

1) The "yes" author would be excited, and would no longer conduct sex difference research
2) The "yes" author would be surprised, because men and women can't be treated the same
3) The "yes" author would be supportive, but would still encourage sex difference research
4) The "yes" author would be disappointed, because sex difference research would be ignored

APPLY: The "no" author would most likely support which of these child-rearing techniques?

1) Fostering an understanding of how boys and girls are similar
2) Treating boys and girls completely equal in every way possible
3) Fostering an understanding of how boys and girls are different
4) Fostering an understanding of both similarities and differences

ANALYZE: Which author would agree with the following statement? "Sex differences are due to nature, not nurture."

1) Neither author
2) The "no" author
3) The "yes" author
4) Both authors

EVALUATE: The "no" author would agree with which statement?

1) A gender difference ideology is a social construction, promoted by male scientists
2) A gender difference ideology ignores the influence of culture, context, and history on women
3) A gender difference ideology is destructive and must be eliminated from our society

4) A gender difference ideology only serves to improve the position of men, not women

CREATE: If the U.S. government funded an increase in sex difference research,
1) The “no” author would argue that the government is providing a service to men
2) The “no” author would argue that the government is providing a service to politicians
3) The “no” author would argue that the government is providing a disservice to society
4) The “no” author would argue that the government is providing a disservice to children

Was Abraham Lincoln America’s Greatest President?

Rephrased Concept Questions

According to the “yes” author, which is one reason why Lincoln was unfit to be President?
1) He was too controversial and unpopular
2) He lacked experience as a war general
3) He was too arrogant and demanding
4) He lacked administrative experience

What did Lincoln do in order to pass the emancipation amendment?
1) He included some restrictions on slaves, which benefited the North
2) He persisted until members of Congress agreed it was the right thing to do
3) He insisted upon equal rights under the Declaration of Independence
4) He promoted the amendment as the only way to end the Civil War

According to the “yes” author, why was emancipation one of Lincoln’s goals?
1) He wanted to demonstrate the power of his influence
2) He wanted to guarantee the right of self-government
3) He wanted to remain consistent with his moral convictions
4) He wanted to guarantee equal rights to all people

In order to pass the emancipation amendment, Lincoln
1) Provided jobs for relatives of congressmen
2) Agreed to support a bill he previously opposed
3) Provided additional money to Republicans
4) Made exceptions for certain slave owners

Ultimately slavery was outlawed,
1) With overwhelming support from Congress

2) And Lincoln was solely responsible for this outcome
3) Although the situation required compromise
4) Which Lincoln always expected to accomplish

According to the "no" author, Lincoln called for heavy taxes in order to
1) Distribute income from the rich to the poor
2) Help the U.S. economy during wartime
3) Pay for military resources and weapons
4) Provide subsidies for big corporations

The "yes" author supported Lincoln, largely because
1) Lincoln was a warrior for the American dream
2) Lincoln passed the emancipation amendment
3) Lincoln was an effective general and ended the war
4) Lincoln overcame adversity and depression

The "no" author opposes Lincoln, largely because
1) Lincoln supported big corporations
2) Lincoln touted his agenda as morally superior
3) Lincoln was a power hungry dictator
4) Lincoln was dishonest and corrupt

Rephrased Higher Order Questions

APPLY: Lincoln might have used which strategy to pass a bill after it was rejected from Congress?
1) He would have appealed to only those affected by the bill
2) He would have revised and resent the bill
3) He would have appealed directly to all voters
4) He would have appealed to members of Congress

ANALYZE: Which author would agree with the following statement? "Politics sometimes requires that we turn a blind eye."
1) The "no" author
2) Both authors
3) The "yes" author
4) Neither author

EVALUATE: The "yes" author would agree with which statement?
1) Lincoln often followed his convictions and he accomplished a great deal
2) Considering Lincoln's beliefs about emancipation, he conquered an uphill battle
3) Considering Lincoln's depression, his accomplishments are impressive
4) Lincoln could persuade others, though he was unwilling to be persuaded

CREATE: Lincoln would most likely support which proposal if he were alive today?

1) An end to overseas wars, because he had an aversion to bloodshed and violence
2) Limitation of government mandates, because he supported the right to self-government
3) Equal rights for homosexuals, because he supported equal rights for slaves
4) Bailouts for car companies and banks, because he supported corporations in the past

APPLY: Based on Lincoln's handling of the Civil War, he most likely would have

1) Encouraged George W. Bush to negotiate with international leaders following September 11th, 2001
2) Encouraged George W. Bush to seek Congressional approval for the use of federal powers following September 11th, 2001
3) Supported George W. Bush's use of federal powers following September 11th, 2001
4) Disagreed with George W. Bush's use of federal powers following September 11th, 2001

ANALYZE: Which author would agree with the following statement? "The way a situation ended was not always the way Lincoln had wanted it to end."

1) Both authors
2) Neither author
3) The "no" author
4) The "yes" author

EVALUATE: The "no" author would agree with which statement?

1) Lincoln was a cheating, greedy President, but he was responsible for some positive results
2) Lincoln could not be trusted, as he deceived Americans in every situation and at every turn
3) Even though Lincoln did what was best only for the North, he still deserves praise for emancipating slaves
4) Lincoln was responsible for many poor decisions, and America was worse off because of Lincoln

CREATE: Lincoln would most likely support which leadership position if he were alive today?

1) The obligation to always do what is morally right
2) The responsibility to follow the Constitution literally
3) The commitment to protect the American people
4) The necessary use of compromise in politics

Is the Superfund Program Successfully Protecting the Environment from Hazardous Wastes?

Rephrased Concept Questions

When the Superfund program first started, how did the EPA assess future health hazards?

1) They used a risk-based hierarchy system
2) They used projected exposure patterns
3) They assumed future commercial use at the site
4) They assumed future residential use at the site

When EPA inspectors identify a low-risk site, what type of remedy solution might they use?

1) Incineration of waste and toxins
2) Prohibition of drinking water wells
3) Soil vapor extraction
4) Removal of contaminated soil

According to the "yes" author, the EPA has made recent progress in which two areas?

1) Risk assessments and treatment strategies
2) Scientific research and treatment strategies
3) Risk assessments and cleanup effectiveness
4) Scientific research and cleanup effectiveness

Which cleanup strategy was used at the beginning of the Superfund program?

1) Acid water containment
2) Source control treatment
3) Excavation and disposal
4) Soil vapor extraction

According to the "no" author, the Superfund program primarily has to deal with which of the following obstacles?

1) Lack of technology and research
2) Conflicts with other environmental agencies
3) Political opposition to cleanup efforts
4) Resident opposition to cleanup efforts

Tar Creek contamination is spreading because

1) EPA failed to truck out toxic dirt
2) EPA failed to invest money in the cleanup effort
3) EPA failed to improve the air quality
4) EPA failed to contain gallons of acid water

Contamination at Tar Creek was created by
1) An abundance of mine shaft drilling
2) An abundance of harmful radiation use
3) An abundance of toxic waste dumping
4) An abundance of noxious chemical use

If implementation risks are not considered, then
1) Cleanup will become impractical
2) Some techniques will be too costly
3) Superfund sites will be ignored
4) Contamination will increase and spread

Rephrased Higher Order Questions

APPLY: Based on the Superfund risk-based hierarchy system, for a health risk of 5 in 10,000,000 at a cleanup site, which remedy solution should the EPA use?
1) Prohibiting land/water use
2) Removal of waste
3) Treatment of waste
4) Restricting land/water use

ANALYZE: Which author would agree with the following statement? "Waste site cleanup is imperative, therefore we should encourage Congress to increase federal funding."
1) Both authors
2) Neither author
3) The "no" author
4) The "yes" author

EVALUATE: The "yes" author would agree with which statement?
1) The Superfund program needs a lot of improvement
2) The Superfund has made great progress recently
3) The Superfund program has always worked really well
4) The Superfund program is both effective and ineffective

CREATE: Which effort do you predict the "yes" author would support, if there were a huge increase in funds for the Superfund program?
1) An increase in research conducted at sites
2) Greater development of incineration techniques
3) Greater development of flexible treatment options
4) An increase in waste removal programs

APPLY: What would happen if all the pollution were removed from the homes and schools in Tar Creek?
1) Tar Creek residents' would no longer require kidney dialysis

2) The other waste problems in Tar Creek would be fixed
3) The "no" author would be satisfied with the newly cleaned Tar Creek
4) There would still be long-term health consequences from the waste

ANALYZE: Which author would agree with the following statement? "EPA buyouts and settlements increase the amount of money in the Superfund program, but they don't solve the problem."
1) The "yes" author
2) Neither author
3) The "no" author
4) Both authors

EVALUATE: The "no" author would agree with which statement?
1) The Superfund program has always been ineffective
2) The Superfund program needs a lot of improvement
3) The Superfund program is both effective and ineffective
4) The Superfund has made great progress recently

CREATE: How do you predict the "no" author would react if the Superfund program received a substantial amount of increased funding?
1) Fine, but most of the money should go toward cleaning up Tar Creek
2) Good, but Superfund may still be ineffective because of partisan politics
3) Useless, money won't make Superfund more effective
4) Great, budget problems are the main reason for Superfund's ineffectiveness

Did World War II Liberate American Women?

Rephrased Concept Questions

According to the "yes" author, the women's liberation movement began because of
1) Magazine articles promoting a transformation
2) New expectations from husbands and other men
3) The desire of women to test and prove themselves
4) Insistence from feminists that women change

According to the "yes" author, earning money had a large influence on the women's liberation movement because
1) It gave women the confidence to demand greater authority
2) It gave women the independence to challenge their husbands
3) It gave women the opportunity to leave the household
4) It gave women the means to make spending decisions

The “yes” author argues that a delay between idea formation and idea implementation is

1) Frustrating
2) Typical
3) Unexpected
4) Unacceptable

The “yes” author believes that women’s wartime experience played a vital role in a liberation movement, largely because of

1) The wartime need for women to earn money for their family
2) The wartime need for women to fulfill both roles of mother and father
3) The wartime need for women to serve as heads of households
4) The wartime need for women to undertake employment duties

According to the “no” author, few women were able to keep their jobs after the war because

1) Women were encouraged to give their jobs back to men when they returned
2) Women were asked to focus on their family, instead of work, when men returned
3) Women failed to keep up the same workplace efficiency as men were able to
4) Women failed to receive any encouragement from men to continue working

According to the “no” author, single women were encouraged to find husbands after the war because

1) Society did not want them to enter the workforce
2) Society had doubts about single women’s motives
3) Society wanted to encourage a “baby boom”
4) Society sought to keep them out of the workforce

The “no” author argues that wartime experience was responsible for

1) Patriotism in women
2) Responsibility in women
3) Ambivalence in women
4) Dependency in women

The “no” author disagrees that women’s wartime experience contributed to a liberation movement, largely because

1) Popular culture recruited women into the labor force temporarily
2) Popular culture returned its focus to men when they came home
3) Popular culture encouraged women to retain their femininity
4) Popular culture continued to reinforce traditional gender roles

Rephrased Higher Order Questions

APPLY: Regarding women who care for children full-time vs. women who work full-time,

1) The “yes” author would support both types of women for asserting their preference and choosing their own lifestyle
2) The “yes” author would support women who work full-time because they are financially independent
3) The “yes” author would support women who split their time between the two and achieve a family-work balance
4) The “yes” author would support women who care for children full-time because raising a family is an important role

ANALYZE: Which author would agree with the following statement? “World War II modernized society and initiated the revolution of traditional gender roles.”

1) Both authors
2) Neither author
3) The “yes” author
4) The “no” author

EVALUATE: The “yes” author would agree with which statement?

1) Gender roles have changed because of an increasing societal value of equality
2) Gender roles have changed because mothers have influenced their daughters
3) Gender roles have changed because of an accumulation of similar experiences
4) Gender roles have changed because individuals pushed for social transformation

CREATE: Currently, 53% of the workforce is male, and 47% of the workforce is female. The “yes” author would most likely feel that

1) Progress has reached its peak, the current situation is as good as possible
2) Progress is too slow, hiring practices and incentives for women must improve
3) Progress is better than expected, demonstrating the large influence of WWII
4) Progress takes time, but at least the situation is better than it was after WWII

APPLY: When men returned from the war, the “no” author would argue that

1) Men returned to the same social climate at both home and work
2) Men returned to a different social climate at home, but not at work
3) Men returned to the same social climate at home, but not at work
4) Men returned to different social climates at both home and work

ANALYZE: Which author would agree with the following statement? "Women's wartime experience greatly affected their children."

1) The "yes" author
2) Neither author
3) The "no" author
4) Both authors

EVALUATE: The "no" author would agree with which statement?

1) Although women were recruited to join the workforce, politicians urged them to maintain their household duties simultaneously
2) Although women had the opportunity to break out of traditional gender roles, they failed to take advantage of the situation
3) Although women were eager to undertake new roles, society was not prepared to allow women to do so, at least not permanently
4) Although women helped maintain the wartime economy by working, women appropriately returned to their household duties

CREATE: Regarding a present-day military draft of men, the "no" author would argue that

1) Women's wartime roles would be viewed as gender-neutral because wartime does not present the same economic difficulties
2) Women's wartime roles would still be viewed as gender-specific because society today is the same as it was in the past
3) Women's wartime roles would still be viewed as gender-specific because wartime always limits the options of women
4) Women's wartime roles would be viewed as gender-neutral because society is more gender-neutral today than in the past

Appendix E

Textbook Chapters Used in Experiment 3

Chapters and test questions used in Experiment 3 were adapted from the 6th grade Social Studies textbook used at Columbia Middle School in Columbia, Illinois.

Banks, J. A., Beyer, B. K., Contreras, G., Craven, J., Ladson-Billings, G., McFarland, M. A., & Parker, W. C. (1997). *World: Adventures in time and place*. New York, NY: Macmillan/McGraw-Hill.

The Russian Revolution

Vocabulary	*People*	*Places*
Russian Revolution	Alexander II	Russia
tsar	Nicholas II	St. Petersburg
strike	Vladimir Ilyich Lenin	Moscow
communism	Josef Stalin	Soviet Union
totalitarian		

Read Aloud

"Peace! Land! Bread!" This slogan summed up what ordinary Russians wanted most in the bloody, food-starved days of World War I. One group promised to give them all these things and more. Once in power, this group would transform Russia and affect the whole world.

The Big Picture

World War I was the peak of a long era of conflict and revolution. You have already read about political and industrial revolutions that rocked the world in the 1700s and 1800s. In the early 1900s, while the "Great War" still raged, yet another revolution broke out – in **Russia**. The **Russian Revolution** was an extremely important event in modern world history.

In 1900 the Russian empire stretched across parts of Europe and Asia. It included people of many different cultures. Most, though, lived in western Russia, where the land was better suited for the empire's main activity – farming. Most Russians were Christians. Muslims also lived in the empire, however, as well as many Jews.

World leaders took notice when revolutionaries overthrew Russia's leaders in 1917. Revolutionary leaders began to build a government around the ideas of Karl Marx, whom you read about in Chapter 17. The world watched and waited. What would happen in Russia? Would Russia continue to fight in World War I? How would the revolution affect other nations?

Russia Under the Tsars

In the middle 1800s Russia was far from being a world power. While industry changed many parts of Europe, most Russians lived much as they had during the Middle Ages.

At the top of Russia's social pyramid was the **tsar** (ZAHR), or emperor. The tsar ruled with an iron hand. Anyone who displeased the tsar might be killed or sent to prison in Siberia. Find this frozen steppe region on the map.

Beneath the tsar were a handful of rich noble families. At the bottom of Russia's social pyramid were millions of poor farmers. Their crops fed the empire.

Russian Serfs

By the late 1700s France and other European countries no longer had serfs, or farmers, bound to the land. In the early 1800s, however, most Russians were still serfs. Russian law said serfs were the property of their owners, although serfs could not be sold.

By the middle 1800s serf revolts in Russia were increasing in number. Tsar **Alexander II** began to fear a revolution. He also wanted to shift Russia's work force away from farming and toward industry. Alexander decided to abolish serfdom in 1861. To abolish means to end a practice. The Tsar said: "It is better to abolish serfdom from above than to wait until the serfs begin to free themselves from below."

In exchange for freedom and small plots of land, the freed serfs had to pay heavy taxes. Paying the taxes was difficult, since many families were given small areas to farm.

Worlds in Conflict

By the late 1800s Russian cities were growing. Hard times in rural areas forced many former serfs to move to the cities in search of work. By the 1890s factories and mills of the Industrial Revolution were springing up in Russia's capital, **St. Petersburg**.

Two Sides of a City

To poor famers St. Petersburg was a new world. They stared in wonder at the grand winter palace of Tsar **Nicolas II**, who began his rule in 1894. Dozens of mansions, churches, theaters, schools, and universities lined the streets of the city. More than one million people lived in St. Petersburg.

The city also had a less spectacular side. Away from the palace and other beautiful mansions, mills and factories clustered together. Smoke from their chimneys filled the air above the overcrowded apartment buildings where workers lived.

Workers Protest

Inside the factories and mills, conditions were often grim and workers were angry. A protest in 1897 won them a shortened work day – to 11.5 hours.

Factory workers protested again in 1905, shutting down the city with their **strikes**. A strike is a refusal to work in protest of unfair treatment.

On Sunday, January 22, 1905, thousands of striking workers marched toward the Winter Palace to speak with the tsar himself. The tsar's soldiers responded by shooting into the crowd. More than 100 people were killed. Many others were injured. The day became known as "Bloody Sunday."

A storm of revolts and strikes swept through the country after "Bloody Sunday." Tsar Nicholas II agreed to share some of his power with a new elected parliament, called the Duma. The Duma called for changes that would advance democracy and help the poor. The tsar refused. During the next nine years, Nicholas and the Duma were in constant conflict.

War and Hunger

In the years following "Bloody Sunday," unrest deepened in Russia. Things became even worse during World War I. More than a million Russian troops died on the battlefront. Some never even had guns or bullets to protect themselves, since weapons were in short supply. Most of the nation's railroads carried supplies to battle. Only a few trains were available to bring food and fuel to cities. As a result, factories and stores often closed. Many people were left without work. Goods that were already hard to get became even more scarce.

March of 1917 began as one of the coldest, snowiest months that many people in St. Petersburg could remember. The weather kept farmers and their food carts away from city markets. Within the city hungry workers lined up in the cold for hours. They hoped to spend what little money they had on small loaves of bread.

Revolution Begins

The skies cleared and the weather changed in time for a protest held by thousands of unhappy people. For four days, demonstrators jammed the streets of St. Petersburg. Shouts of "Down with the war!" and "Down with the government!" soon drowned out the simple cry for "Bread!"

The tsar's police called for help from soldiers who were staying in the city. Most of the soldiers, however, joined in the protest and turned on the police. With the soldiers' help the protest became a full-scale revolution against the government.

Tsar Nicholas, who was away meeting with his generals, had no idea of what was happening in his capital. By the time he set out to return home, the spirit of revolution had spread. Angry railroad workers forced his train to a standstill. On March 15, 1917, Nicholas II was forced to give up his role as tsar. Sixteen months later he, his wife Alexandra, and their children were executed. The rule of Russian tsars had come to an end. Who would rule the giant nation now?

A New Government

After the revolution in March, the Duma chose leaders to run the country. Russia's many problems, however, continued. World War I was still underway

and Russian military leaders demanded that their troops be withdrawn from the front. City workers went on striking in protest of even longer bread lines and lower wages. Many farmers, hungry and impatient for change, began seizing land for themselves.

Meanwhile a political group called the Bolsheviks was gaining strength. The Bolsheviks were led by a Russian lawyer named **Vladimir Ilyich Lenin** (VLAD uh meer IHL yitch LEN in). He believed that a different kind of revolution was necessary to change the government. The Bolsheviks planned a socialist revolution based on the ideas of Karl Marx whom you read about in Chapter 17. They wanted workers to control the government and own all property. Lenin promised Russians "Peace, Land, and Bread."

The Bolsheviks Take Control

With the support of the soldiers in St. Petersburg, Lenin and the Bolsheviks overthrew the Duma in November 1917. Soon after this second revolution they pulled Russian troops out of the Allied war effort. Russia began peace talks with Germany. The Bolsheviks allowed workers to control factories and farmers to use the farmland of wealthy nobles. The Bolsheviks also moved the capital of Russia south to the ancient city of **Moscow**.

The new Bolshevik government had many opponents. Landowners, factory owners, and nobles were losing their rights, as well as their wealth and power. Christians and different ethnic groups also opposed the government. These people led a civil war against Lenin and the Bolsheviks.

The Russian people were already battered from world war and revolution. Their suffering became even greater, however, during this new civil war. Between 1918 and 1920, millions died from disease and starvation, as well as in violent battles.

Communism

Lenin wanted to create **communism** in Russia. Communism is a political and economic system in which all land and all businesses are controlled by the government.

In the months before the outbreak of the civil war, Lenin wrote, the Bolsheviks had left "one foot in socialism." In other words, they had been moving slowly toward a society controlled by workers. Now, though, Bolshevik leaders took harsh steps to achieve communism in Russia.

The Bolsheviks outlawed all private property, including farms. Farmers were forced to give all of their grain to the government. Lenin replaced factory workers' committees with new managers who were controlled by the Communist Party. Citizens were called upon to serve in the military. To break people's loyalty to religion, the Bolsheviks closed churches and arrested religious leaders. Lenin insisted that all loyalty be focused on the government.

Union of Soviet Socialist Republics

By 1920 the Bolsheviks had defeated their enemies. Two years later they renamed the old Russian empire. The new nation became known as the Union of

Soviet Socialist Republics, or the **Soviet Union**. The soviets were councils of workers and soldiers formed during the revolution.

In 1922 Lenin became ill. He struggled to return to work, but another leader in the Soviet government was growing more powerful. His name was **Josef Stalin**.

Stalin's Rule of Terror

Lenin died in 1924. Soon after, Josef Stalin became the new leader of the Soviet Union. In 1928 Stalin began working to make the Soviet Union stronger. He drew all power into the government. Stalin also created huge collective farms. Collective farms were run by the government and worked by many families. People not needed on farms were sent to work in mines and factories springing up across the nation. An economy completely controlled by government is called a command economy.

Within just 20 years the Soviet Union became one of the world's strongest industrial nations. Thousands of railroad lines crisscrossed the country, linking towns and cities that had never been connected before. Around 1900 many Russian farmers had never seen a tractor. By the 1940s Soviet factories were making more tractors than any other factories in the world.

Totalitarian Rule

People paid a huge price, however, for growth and change in the Soviet Union. Stalin used **totalitarian** (toh tal ih TAIR ee un) methods to rule the nation. In a totalitarian society, a dictator, often representing a single political party, controls all aspects of people's lives. Stalin and the Communist Party controlled the Soviet Union through fear and terror. For many, life was more difficult than it had been under the tsars. People were arrested for speaking their minds freely or for writing to friends in other countries. Many managers were killed because their factories or farms did not produce an expected amount. Stalin also ordered his secret police to arrest anyone who he thought challenged him in any way.

Many of those arrested were religious leaders. Their followers were forced to worship secretly or face arrest themselves. Stalin had more than 15 million people killed or sent to prison camps in Siberia. Almost half of them were Ukrainians. Many starved because the collective farms failed to produce enough food. Large numbers of people were sent to camps where religious leaders, teachers, workers, and others Stalin considered "enemies of the people" were imprisoned.

Oil, iron, timber – all the resources of the Soviet Union's new industry – were in great supply in Siberia. Since few people lived there, Stalin used political prisoners to help collect the resources.

One women's camp had the job of cutting down trees. One of the prisoners, a teacher, described the camp this way:

The cold and the hunger; the hunger and the cold. This must have been the blackest, the most [deadly], the most evil of all my winters in the camps.

Why It Matters

In the early 1900s life changed dramatically in Russia during a period of revolution. Many of the changes that took place became the foundation of a communist system of government. For this reason, the Russian Revolution is also known as a communist revolution. One of the revolution's many effects was the formation of the Soviet Union.

Revolutionary leaders had promised "peace, land, and bread." Under the communist government, however, most people in the Soviet Union had none of these things. Millions were killed and sent to prison camps in Siberia by Josef Stalin. Stalin used totalitarian methods to rule the nation.

For many, suffering worsened when the Soviet Union and many other countries became involved in another world conflict. To the west of the Soviet Union, a dictator in Germany was making plans that would lead to war.

Sum It Up

- Millions of serfs under Russia's tsars lived in poverty. The abolition of serfdom in 1861 gave farmers a limited amount of freedom.
- The Russian Revolution began in 1917 as a revolt against World War I, the tsar, and poor working and living conditions. Seven months later Lenin and the Bolsheviks seized control, bringing communism to the country they later renamed the Soviet Union.
- Under Stalin, the Communist Party controlled the Soviet Union using totalitarian methods.

Think About It

1. What were the policies of the Soviet Union regarding religious beliefs and practices?
2. Define the term *communism*.
3. FOCUS: How were the governments led by Tsar Nicholas II and Josef Stalin similar? How were they different?
4. THINKING SKILL: Describe Josef Stalin's *point of view* about the need to totally control the economy of the Soviet Union.
5. GEOGRAPHY: Why might Stalin have chosen Siberia as a site for prisons?

World War II

Vocabulary
fascism
inflation
depression
propaganda
World War II
Axis
Allies
concentration camp
Holocaust

People
Adolf Hitler
Winston Churchill
Franklin Roosevelt
Anne Frank

Places
Pearl Harbor
Normandy

Read Aloud

"I pray to the Almighty God that He shall spare the nations the terrible sufferings that have just been [forced] on my people… Are [you] going to set up the terrible precedent of bowing before force?"

In 1936 Ethiopian emperor Haile Selassie (H**ī** lee suh LAS ee) appeared before the League of Nations to protest Italy's invasion of his African country. The League, however, did not come to Selassie's aid. Ethiopia would not regain its independence for nearly five years. During much of that time, the world was once again plunged into war.

The Big Picture

After the signing of the Treaty of Versailles, the countries that had fought in World War I turned to their own affairs. In the last lesson you read about the communist revolution in Russia. In 1922, the dictator Benito Mussolini and his Fascist (FASH ihst) party rose to power in Italy.

The Fascists believed in a powerful leader, totalitarian government, and an extreme form of nationalism. They supported a government whose goals they thought to be more important than those of individual people. This type of government came to be known as **fascism**. In some places fascism also came to mean hatred of certain ethnic groups.

After Italy took control of Ethiopia in 1936, Mussolini joined forces with another fascist dictator, **Adolf Hitler** of Germany. The people of nearby nations began to see that fascism was a serious threat to peace.

Germany After World War I

In 1919 Germany began to live by the conditions of the Treaty of Versailles. The treaty stripped Germany of land and forced it to pay huge fines.

To meet these expenses the German government began printing large amounts of paper money. Before long Germany had printed so much money that it began to lose its value. The result was a period of inflation, or rising prices. Huge amounts of money were needed even to buy necessities such as food. By 1923 inflation had made German money practically worthless, and people's savings were gone.

In that year a bitter ex-soldier named Adolf Hitler led an attack against the German government in the state of Bavaria. Although the attack failed and Hitler

was jailed, many Germans supported his actions. His followers were known as the Nazi (NAHT see) party.

Fascism in Germany

By the early 1930s, Germany and much of the world suffered a **depression** (di PRESH un). During a depression, fewer goods are produced, prices drop, many people lose their jobs, and money is hard to get.

During these hard times Hitler used **propaganda** (prahp uh GAN duh) to convince Germans that their nation would once again become powerful. Propaganda is the spreading of certain ideas or attitudes that have been exaggerated or falsified to advance a particular cause.

Hitler's propaganda spread the false idea that the Germans were a "master race," meant to rule the world. The Nazis wrongfully blamed Germany's Jews, along with the Treaty of Versailles, for the depression that was devastating the country. Promising to raise Germany back to glory, Hitler once again tried to gain control in 1933. This time he succeeded.

Hitler ruled as a fascist dictator, forming an alliance with Mussolini in Italy. He and the Nazis stirred up hatred against Jews. In five years the Nazis' plans would lead to the largest war in history.

A Second World War

In 1938 Hitler ordered Nazi troops to occupy neighboring Austria. With this command, Hitler knowingly broke the rules of the Treaty of Versailles. Then, in March 1939, Hitler seized control of Czechoslovakia. After years of trying to avoid war with Germany, the leaders of Britain and France promised to defend Hitler's next target – Poland. Europe was on the brink of war once again.

The German Advance

World War II began in Europe on September 1, 1939. On that day German tanks began a *blitzkrieg* (BLIHTZ kreeg), or "lightning war," in Poland. Hitler and Josef Stalin, whom you read about in Lesson 2, had recently signed a friendship treaty. With the help of the Soviet Union, Germany defeated Poland within weeks. Britain and France declared war on Germany but had not been able to defend their ally, Poland.

Eight months later German forces turned west. Hitler's armies quickly overran Belgium. They went on to seize Paris by June 1940. Hitler's fighting method of blitzkrieg was proving very effective. Germany had beaten France – a major world power – in oly six weeks! With much of France under German control, Hitler made Britain the next Nazi target.

The Battle of Britain

The British people prepared for the worst. The country's leader, Prime Minister **Winston Churchill**, declared:

> *We shall fight on the seas and oceans, we shall fight with growing confidence and growing strength in the air, we shall defend our island, whatever the cost may be…. We shall never surrender.*

Two months later, in August 1940, the Battle of Britain began. For almost a year German planes bombed the island nation every night. The British air force fought back. Although many sought safety in underground shelters, more than 12,000 British people were killed in the fighting. Despite the cost Britain did not surrender. The nation stood firm, as Churchill had predicted.

Weather Plays a Part

In June 1941 Hitler ended the bombing of Britain. Germany had lost more than 2,000 planes, along with their crews. Having failed in Britain, Hitler decided to break his treaty with Stalin. He ordered his armies to turn east and invade the Soviet Union. The Germans began what was to become a three-year struggle for control of major Soviet cities and supply centers. Millions of Soviet soldiers and civilians died during the struggle.

The Soviet Union now became an ally of Britain. In spite of their political differences, the British welcomed the Soviets in the fight against their common enemy, the Nazis. By November 1941 German troops were very close to one of their goals: the Soviet capital, Moscow. Soviet armies fought to defend their capital and their country. The German troops were finally stopped, however, by a deadly northern winter. On December 6, the near-frozen Germans began to retreat. It would not be the last time nature played a part in the outcome of the war.

An Attack on the United States

War had begun earlier in Asia than it had in Europe. Japan had hoped to create an empire with an endless supply of raw materials and labor for industry. By 1931 Japanese forces had invaded northern China. Later Japan conquered about one quarter of China and some islands off the coast of South Asia. Find the region of Japanese expansion on the map on pages 544-545.

In 1940 Japan formed an alliance with Germany. The conquests and the alliance created tension between Japan and the United States, which was against Japan's continuing expansionist policy. Japan was determined to stop the United States from involvement In its expansionist plans.

On December 7, 1941, Japan launched an attack without any warning or declaration of war. The target was the United States naval base at **Pear Harbor**, Hawaii. More than 2,000 people died in the attack. The United States was no involved in World War II.

President **Franklin Roosevelt** declared war on Japan on December 8, 1941. Three days later, on December 11, Germany and Italy declared war on the United States. Japan, Germany, Italy, and their other allies were known as the **Axis**. The **Allies** included Britain, France, the Soviet Union, the United States, and China, among others. The United States had to fight Japan in Asia and Germany and Italy in Europe and Africa. As in World War I, United States forces would be very important to the Allied war effort.

"The Longest Day"

For three years the United States, Britain, and other Allies fought the Nazis in Europe and North Africa. In that time, the Soviet Union struggled to push back and destroy the German invaders on its soil. Finally, Allied leaders prepared to put a risky plan into action. On the night of June 5, 1944, the Allies would begin a surprise invasion of Axis-held France. If they succeeded, Germany would be surrounded on three sides – west, east, and south. The allies' code name for this operation was D-Day.

Allied leaders prepared their forces to land on the beaches of **Normandy**, France, at dawn on June 6. They would reach shore while the tide was low so that German weapons on the beach would be open to attack. Months earlier, weather experts had concluded that the best conditions for an attack would exist between June 5 and 7. One June 4, though, a terrible storm raged across the English Channel. Would nature stop the biggest sea invasion in history?

The storm actually helped the Allies. German commander Erwin Rommel believed that the Allies would not invade during such weather. He traveled home to Germany for a few days, just when D-Day arrived. The Allies attacked. Over 11,000 Allied planes dropped bombs and over 2,700 ships unloaded almost 200,000 men onto the beaches of Normandy. Find Normandy on the map.

Afterwards, an Allied soldier said D-Day seemed like "the longest day" of his life. At the end of that day, allied forces held the beaches. The allies would now begin to push the Axis powers east across Europe and west from the Soviet Union.

The End of the War

Less than a year after D-Day, Allied forces closed in around Germany. With the Soviet army already in the German capital of Berlin, Adolf Hitler killed himself to avoid capture on April 30, 1945. One week later, on May 7, 1945, Germany surrendered. Japan's leaders, however, refused to give up the struggle for power.

United States leaders considered using a newly developed bomb against the Japanese. Invading Japan could lead to many deaths on both sides. Could the tremendously powerful atomic bomb bring about Japan's surrender? On August 6, 1945, the United States dropped the first atomic bomb ever used in warfare on the Japanese city of Hiroshima (hihr uh SHREE muh). Most of the city was destroyed in seconds, and at least 80,000 people died.

Japan did not surrender. Three days later the United States dropped another atomic bomb on the city of Nagasaki (nah guh SAH kee). Japan surrendered on August 14, 1945. The most terrible war in history was finally over.

The Terrible Effects of Fascism

In the days before their defeat, German and Japanese commanders rushed to hide evidence of their **concentration camps**. Concentration camps are places where people are imprisoned because of their heritage, religious beliefs, or political views. Prisoners in Japanese and Nazi concentration camps were tortured and often killed. Millions of others were murdered as well.

The Nazis murdered about 6 million Jews, or two-thirds of Europe's Jewish population, in concentration camps or by execution squads. These people including women, children, and elderly people, had committed no crime. They were not soldiers. They were killed for no other reason than that they were Jewish. This deliberate destruction of human life is called the **Holocaust** (HOL uh kawst). About another 6 million people, among them Gypsies, Poles, Russians, and Slavs were also murdered in Nazi concentration camps.

One of the millions of young Jews who died in the camps was 15-year-old **Anne Frank**. She and her family spent two years hiding in the Netherlands before Nazi soldiers captured them. What did Anne Frank believe about people and about the future? Do you find her point of view surprising?

Many Voices Primary Source

Excerpt from The Diary of Anne Frank, July 1944.

It's really a wonder that I haven't dropped all my ideals, because they seem so absurd and impossible to carry out. Yet I keep them, because in spite of everything I still believe that people are really good at heart. I simply can't build up my hopes on a foundation consisting of confusion, misery, and death. I see the world gradually being turned into a wilderness, I hear the ever approaching thunder, which will destroy us too, I can feel the sufferings of millions and yet, if I look up into the heavens, I think that it will all come [out] right, that this cruelty too will end, and that peace and tranquility will return again.

Why It Matters

World War II was the largest war in history. Unlike World War I, which had been fought mostly in Europe, World War II took place in Europe, Asia, Africa, and the islands of the Pacific. While many battles took place on land, there were sea battles on the world's oceans, as well. The war left as many as 50 million people dead. Many millions more would be affected by its horrors throughout their lives.

People once again began to adjust to peace after a world war. It was not always easy. Destroyed roads, bridges, homes, and cities around the world had to be rebuilt. There were other serious problems, too.

Leaders of the United States and Western Europe feared the communist government of the Soviet Union. Soon the two most powerful Allies, the United States and the Soviet Union, would become bitter enemies. You will read about their conflict later in the chapter.

Sum It Up

- In the 1930s Nazi leader Adolf Hitler used propaganda to convince many Germans that their nation could return to its former power.
- A world depression in the 1930s caused suffering in many nations and helped to bring about the rise of fascist dictators, such as Hitler.
- The Nazis murdered about 6 million Jews in concentration camps. This became known as the Holocaust. There were also some 6 million other victims of the Holocaust, including Gypsies, Poles, Russians, and Slavs.

- Japan attacked and conquered parts of Southeast Asia and the Pacific. After Japan attacked Pearl Harbor, the United States entered World War II. The war came to an end after the United States used two atomic bombs on Japanese cities.

Think About It

1. How did the United States help the Allied war effort?
2. Why was D-Day an important battle?
3. FOCUS: How did Hitler use the problems created by inflation, the depression, and unemployment to make himself dictator of Germany? How did he use this power to bring about World War II?
4. THINKING SKILL: List three ***facts*** and one ***opinion*** about fascism.
5. GEOGRAPHY: What role did the weather and time of attack play in the planning and outcome of D-Day?

Appendix F

Concept and Higher Order Questions Used in Experiment 3

Note. Correct answers are underlined, and the type of higher order question (apply, analyze, evaluate, or create) is indicated below for illustrative purposes, but correct answers and higher order types were not revealed to subjects during testing.

Russian Revolution

Concept Questions

Why did Alexander II abolish serfdom?
A) To focus Russia's work force toward farming
B) To prevent an uprising from farmers
C) To take away land from farmers
D) To reduce taxes paid by farmers

What happened once Alexander II abolished serfdom?
A) Farmers increased food production
B) Farmers made a lot more money
C) Farmers moved closer to cities to find work
D) Farmers moved to rural areas to find work

Under Alexander II, which area of Russia experienced the greatest growth?
A) Cities
B) Siberia
C) Palaces
D) Farms

What led to Nicholas II's agreement to share some power with the Duma?
A) Communal farms
B) Bloody Sunday
C) Disease and starvation
D) Resistance from nobles

Why were Nicholas II and the Duma in constant conflict?
A) Because Nicholas II wanted to help the poor
B) Because the Duma wanted to support communism
C) Because the Duma wanted to help the poor
D) Because Nicholas II wanted control of all of Russia's power

Why was Nicholas II forced to give up his role as tsar?
E) Because the Duma elected a new tsar

F) Because Stalin took over the government
G) Because his wife and children moved to Moscow
H) Because of angry protestors, soldiers, and railroad workers

At first, who did Lenin believe should control the government?
A) Tsars and nobles
B) People and workers
C) Farmers
D) The Duma

After a few years under Lenin, what happened to Russia's farmers?
A) Farmers had complete control over their farms
B) Farmers had to give all of their grain to the government
C) Farmers were forced to sell their farms
D) Farmers were running out of farm land

Why did Lenin close churches and arrest religious leaders?
A) To focus all loyalty on work and factories
B) To focus all loyalty on farm and food production
C) To focus all loyalty on families and communities
D) To focus all loyalty on the government

Why did Stalin become the new leader of the Soviet Union?
A) Because of civil war
B) Because Alexander II took over
C) Because Lenin died
D) Because of a strike

How did Stalin try to make the Soviet Union stronger?
A) He gave all power to the people
B) He built a lot of railroads and tractors
C) He allowed the Duma to make decisions
D) He improved working conditions in factories

Under Stalin, how would you describe everyday life for the Russian people?
A) People were free to do whatever they wanted
B) Stalin controlled all aspects of people's lives
C) Stalin forced all people to go to church
D) People were allowed to choose their careers

Higher Order Questions

APPLY: Based on what you know about Alexander II, how would he react if his military was about to revolt?
A) He would try to prevent the revolt before it happened

B) He would wait until the revolt started before taking action
C) He would leave Russia and avoid the revolt before it happened
D) He would order his police to attack the military to stop the revolt

ANALYZE: Which person would agree with the following statement? "People must pay taxes in exchange for freedom."
A) Nicholas II
B) Alexander II
C) Lenin
D) Stalin

EVALUATE: Which statement is an accurate summary of Alexander II's views?
A) Stop something bad before it happens
B) The government shouldn't control anything
C) The government should control all power
D) Farming is the key to Russia's success

APPLY: Based on what you know about Nicholas II, how would he treat poor people?
A) He would share some power with the poor
B) He would help the poor
C) He would take money away from the poor
D) He would ignore the poor

ANALYZE: Which person would agree with the following statement? "Revolutions are hard to prevent."
E) Alexander II
F) Lenin
G) Nicholas II
H) Stalin

EVALUATE: Which statement is an accurate summary of Nicholas II's views?
A) A tsar should never share power with anyone
B) Advancing democracy is important
C) Sharing power prevents strikes and revolts
D) Sharing power is sometimes necessary

APPLY: Based on what you know about Lenin, what probably changed his beliefs from socialism to communism?
A) The starvation
B) The civil war
C) World War I
D) The Duma

ANALYZE: Which person would agree with the following statement? "I tried to help the poor, but that upset all the landowners and nobles. You can't make everyone happy."

A) Lenin
B) Nicholas II
C) Alexander II
D) Stalin

EVALUATE: Which statement is an accurate summary of Lenin's views?

A) Ultimately, control by the government was what was best for Russia
B) Ultimately, control by the people was what was best for Russia
C) Ultimately, control by the Duma was what was best for Russia
D) Ultimately, control by the farmers was what was best for Russia

APPLY: Based on what you know about Stalin, how would he have reacted when Alexander II abolished serfdom?

A) Stalin would have agreed, because farmers deserve to have some freedom from the government
B) Stalin would have agreed, because farmers could still be controlled by paying heavy taxes
C) Stalin would have agreed, because it was the right thing to do and it would help everyone
D) Stalin would have agreed, because Russia's workforce should be focused on farming not industry

ANALYZE: Which person would agree with the following statement? "People are the most productive when they are told what to do by one person, instead of listening to many people or doing what they want."

A) Nicholas II
B) Lenin
C) Stalin
D) Alexander II

EVALUATE: Which statement is an accurate summary of Stalin's views?

A) A country will be strongest with a few people in charge
B) A country will be strongest with many people in charge
C) A country will be strongest with all people in charge
D) A country will be strongest with only one person in charge

Concept Questions

What happened to Britain during World War II, when Winston Churchill was Prime Minister?

A) Britain stood firm against attacks by Germany
B) Germany stood firm against attacks by Britain
C) Germany surrendered to Britain
D) Britain surrendered to Germany

What did Franklin Roosevelt do during World War II?

A) He dropped an atomic bomb on Japan
B) He joined the Axis war effort
C) He killed Adolf Hitler
D) He declared war on Japan

Why is Anne Frank inspirational?

A) Because she fought against the Nazis
B) Because she had a positive outlook on life
C) Because she survived the concentration camps
D) Because she helped other Jews

Who did Hitler join forces with?

A) Selassie, the leader of Ethiopia
B) Roosevelt, the leader of the U.S.
C) Churchill, the leader of Britain
D) Mussolini, the leader of Italy

Why were Hitler's armies effective at occupying Poland and France?

A) Because of a large army
B) Because of special weapons
C) Because of a lightning war
D) Because of fast surrenders

Why did Hitler stop attacking the Soviet Union?

A) Because the Soviet Union was too strong to defeat
B) Because the Soviet Union had a winter storm
C) Because the Soviet Union teamed up with Italy
D) Because the Soviet Union and Germany signed a treaty

Why did Hitler join forces with Japan?

A) So they could both take over the United States
B) So they could work together to expand their empires
C) So Germany could build an army base in Japan
D) So Japan wouldn't join the Allied Forces

Why did Germany surrender at the end of World War II?
A) Because of Hitler's death after D-Day
B) Because of the atomic bomb
C) Because the U.S. attacked Germany
D) Because Germany's army ran out of resources

How did Hitler gain more followers for his Nazi party?
A) He promised to improve Germany's living conditions
B) He promised to end Germany's economic depression
C) He promised to expand Germany's industry jobs
D) He promised to bring back Germany's glory and power

Under Hitler, which goals were the most important to achieve?
A) The people's goals
B) Japan's goals
C) The government's goals
D) The Allies's goals

Why did Hitler invade the Soviet Union?
A) Because he failed during the Battle of Britain
B) Because he wanted to break the Treaty of Versailles
C) Because he wanted to steal Russia's military
D) Because he failed during the Battle of Normandy

Under Hitler, what led to the largest number of human deaths?
A) The attack on Britain
B) The attack on Poland
C) Concentration camps
D) Atomic bombs in Japan

Higher Order Questions

APPLY: Based on what you know about Franklin Roosevelt, what would he do if Spain attacked the U.S.?
A) He would surrender to Spain
B) He would drop an atomic bomb on Spain
C) He would attack Spain in return
D) He would negotiate with Spain

ANALYZE: Which person would agree with the following statement? "Even if people do bad things, they are still good people."
A) Franklin Roosevelt
B) Winston Churchill
C) Adolf Hitler

D) Anne Frank

EVALUATE: Which statement is an accurate summary of Winston Churchill's views?
A) War is sometimes necessary, but not always
B) A country should always protect its soldiers
C) Don't give up, even when it's tough
D) Help from other countries is the only way to win

APPLY: Based on what you know about Hitler, how might he have reacted if Japan didn't join forces with Germany?
A) He would have ignored Japan
B) He would have attacked Japan
C) He would have protected Japan
D) He would have signed a treaty with Japan

ANALYZE: Which person would agree with the following statement? "The Treaty of Versailles was a bad idea."
A) Franklin Roosevelt
B) Anne Frank
C) Adolf Hitler
D) Winston Churchill

EVALUATE: Which statement is an accurate summary of Hitler's views?
A) The German military was responsible for Germany's depression
B) Some German religious groups were responsible for Germany's depression
C) All German people were responsible for Germany's depression
D) The German government was responsible for Germany's depression

APPLY: Based on what you know about Hitler, how would he try to defeat the United States?
A) He would use a propaganda strategy
B) He would use a fascist strategy
C) He would use an economic strategy
D) He would use a blitzkrieg strategy

ANALYZE: Which person would agree with the following statement? "Building a powerful country is very important, regardless of the cost."
A) Adolf Hitler
B) Franklin Roosevelt
C) Haile Selassie
D) Winston Churchill

EVALUATE: Which statement is an accurate summary of Hitler's views?
A) In order to build a master race, Germany must expand its empire

B) In order to build a master race, Germany must make more money
C) In order to build a master race, Germany must protect only German people
D) In order to build a master race, Germany must increase the number of jobs

APPLY: Based on what you know about Hitler, how would he have reacted if he was alive when Germany surrendered?
A) He would have refused to surrender
B) He would have been glad to surrender
C) He would have been sad to surrender
D) He would have agreed to surrender

ANALYZE: Which person would agree with the following statement? "Loyalty to one's country is more important than any other type of loyalty."
A) Franklin Roosevelt
B) Haile Selassie
C) Winston Churchill
D) Adolf Hitler

EVALUATE: Which statement is an accurate summary of Hitler's views?
A) By invading the Soviet Union, Germany can create a master race
B) By invading the Soviet Union, Germany can expand its empire
C) By invading the Soviet Union, Germany can increase food production
D) By invading the Soviet Union, Germany can strengthen its military

Appendix G

Initial Quiz and Delayed Test Performance (Proportion Correct) as a Function of Learning Condition for All Subjects in Experiment 3

	Pre-Quiz	Post-Quiz	Review Quiz	Final Concept Test	Final Higher Order Test	Delayed Average
Non-Quizzed				.62 (.18)	.55 (.19)	.59
Higher Order Quizzes (3X)	.37 (.17)	.67 (.20)	.80 (.18)	.63 (.19)	.74 (.21)	.69
Mixed Quizzes (3X)	.38 (.18)	.71 (.20)	.86 (.16)	.89 (.17)	.80 (.21)	.85
Average	.38	.69	.84	.71	.70	

Note. Standard deviations are displayed in parentheses. N = 122 (all present and absent student, not including special education students), although the number of subjects in each cell varies between n = 116 to n = 121. Analyses of variance and *t*-tests with all subjects revealed the same pattern as reported in Experiment 3, pp. 68-71.