2017

The Influence of Testing and Content Presentation Method on Mandatory Federal Employee Training

Corey Patrick Boswell
University of South Carolina

The Influence of Testing and Content Presentation Method on Mandatory Federal Employee Training

by

Corey Patrick Boswell

Bachelor of Arts
University of South Carolina, 2006

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Educational Psychology and Research

College of Education

University of South Carolina

2017

Accepted by:

Kelly Lynn Mulvey, Major Professor

Matt Irvin, Committee Member

Robert Johnson, Committee Member

Kellah Edens, Committee Member

Stuart Senter, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Dedication

Above all, the purpose of this labor is to glorify God. It is first to His majesty that I dedicate this work. This project was made possible through His guidance and wisdom, and for that I stand in awe. Secondly, I owe an incalculable debt to my family for providing both support and inspiration in equal measure throughout this lengthy process. My wife, Mynahgi, and my son, Dalton, have shown patience and love, even when work outweighed family time. It is my hope that the effort undertaken for this project will directly lead to a career where I have the professional autonomy to divide my availability in a way that favors my responsibilities as a husband and father. Finally, to my parents, Charlie and Harriette. They are gone from this world, but the lessons they taught me endure. From them I learned the importance of faith, strength, perseverance, selflessness, compassion, and laughter.

Acknowledgements

I want to take this opportunity to thank those individuals who contributed to the current project. Beginning with my dissertation committee: Dr. Kelly Lynn Mulvey, Dr. Matt Irvin, Dr. Robert Johnson, Dr. Kellah Edens, and Dr. Stuart Senter. Your collective expertise, unfailingly timely advice, and constructive encouragement will not be forgotten. Also, I would like to thank Dr. Christine DiStefano and Ms. Wilma Sims for statistical consultation during the data analysis phase, and Dr. Michael Seaman for insightful dialogue regarding experimental design issues. I want to thank everyone at the National Advocacy Center (NAC) who helped transform this study from idea into reality. Dr. Angela Dooley has been an invaluable resource for helping to successfully navigate this project at every stage. Julius Grant was instrumental in making the stimulus design phase a smooth process. The remainder of staff and administration at the NAC offered great support for the project as well. Finally, I would like to thank the participants, who gave of their time and energy without expectation.

Abstract

A substantial amount of evidence exists suggesting that including tests as part of learning events can promote greater long-term retention (known as the "testing effect"). In the current study, the testing effect was analyzed in the context of mandatory federal legal training. The classic information processing perspective provided a theoretical backdrop for the experimental design. Participants (N = 383) received specialized training content through one of three modalities (text-, audio-, or video-based). Additionally, instructional style (test vs. no-test) was manipulated in conjunction with content presentation method. It was predicted that participants would perform better on the final assessment in conditions employing testing as an instructional tool, and that no differences would emerge in performance according to presentation method. No overall main effects of testing or presentation method emerged. However, participants scored better on lower-order items in the video-based presentation, and increased scores on the higher-order questions were linked with text-based presentation. Additionally, in both lower- and higher-order sets of questions, participants scored better if they had previously viewed the content in past training events. Implications and recommendations are discussed subsequently.

Table of Contents

Chapter 1: Introduction

From antiquity onward, it is inarguable that teachers and mentors have spent a great deal of time considering how best to transfer knowledge to their intellectual apprentices. Thorndike and Woodworth's (1901) early work hinted that the "spread of practice" (i.e., "performance benefits") extended only to similar areas of mental functioning mirroring the practice context. More recent research has examined the nature of knowledge transfer from academic to workplace settings (Garraway, Volbrecht, Wicht, & Ximba, 2011). While the results are mixed, clearly the field of learning remains so crucial to the understanding of human behavior to this day that investigators representing many distinct disciplines (such as education and psychology, to name two) regularly reach out across theoretical aisles to collectively focus on the enduring questions that spring from its study. One of the more basic, yet persistent, lines of inquiry frames the question: How can the *efficiency* of learning processes be improved? In brief reply, the main point of this proposal is to suggest that adopting a practice of "test-enhanced" instruction, essentially the inclusion of periodic assessments during learning (Roediger & Karpicke, 2006a), may assist with beginning to develop a more concrete, workable answer. The subsequent discussion will take an information-processing perspective (Simon & Barenfeld, 1969; Schneider & Shiffrin, 1977; Morrison, Burnham, & Morrison, 2015) through multiple learning contexts, ultimately extending an application to mandatory federal employee training.

Currently, little work exists examining the effectiveness of the processes underlying instruction and learning in federal mandatory training in the United States government, though attention to outcome-based evaluation is increasing (Heinrich, 2008). Requisite training for federal employees is a costly enterprise. It is a government "development" service that (through ongoing review) is in serious need of restructuring to address systematic deficiencies, such as prioritizing training foci and comparing content delivery systems (GAO, 2012). Comprehensively restructuring the overall practice with an eye toward austerity could conceivably result in a drastic line item expenditure reduction in the larger federal budget. Some evidence supports the notion that education designated as "non-optional" can produce negative "downstream" consequences. Troublingly, research indicates that an employee's motivation to transfer training-acquired knowledge (that is, apply learned information in a work environment) is substantially minimized when the training episodes are mandated, rather than voluntary (Curado, Henriques, & Ribeiro, 2015). Hung, Sun, and Yu (2015), for example, found that young (2^{nd}-grade) students performed better in conditions representing elevated degrees of challenge when learning the concepts of addition and subtraction. The authors believed that greater challenges invoked higher levels of motivation in the children.

While the current project exclusively employs a *mandatory* training module, a larger purpose of this effort is to identify principles and practices that can generalize to the broader scope of federal training to the greatest extent possible. If federal trainees know they have the option to read a few paragraphs (at their own pace) and take a summative test, as opposed to spending an hour or longer watching a video, they may be more motivated to engage in deeper mental processing of the material.

In practical terms, improving the federal system for training employees across agencies could illuminate observable effects like reducing the waste/misuse of limited resources such as time and tax dollars. Current spending on federal training is in the neighborhood of hundreds of millions of dollars annually (Peterson and McCleery, 2011). Further, overhauling the training system may lead to the intrinsic benefit of heightening employee satisfaction, seen through performance-based indicators like job competency and more favorable evaluations.

Background Theory

Cognitive Processing Theories

Following the recession of behaviorist ideology, which dominated much of the first half of 20th century psychological research, the contrasting *cognitive* perspective usurped the former paradigm in relatively rapid fashion. Central to the principles of the new branch of thinking was the idea of man as "information processor," likened in many ways early on to that of a programmable computer (Simon & Barenfeld, 1969). With the focus primarily trained on how information was taken in, manipulated for storage, and then later retrieved, computer models became an intuitive proxy through which to imagine the inner workings of human cognition. Shortly, the information processing view was applied to the context of learning, including accounts of how multiple representations of knowledge in memory systems may influence forgetting versus remembering information (Andre, 1972).

Refining the characterization of the information processing approach, Schneider and Shiffrin (1977) offered a binary parsing of the (general) theory's functional underpinnings. They claimed it was sensible to view processing as either automatic or

effortful. Dependent on familiarity with content and/or procedural factors, this conceptual structure seems intuitive. Additionally, the information processing view was directly responsible for the expansion of one of psychology's most broadly (and routinely) investigated topics: memory. Baddeley and Hitch (1994) elaborated a multi-component memory model, including verbal and visual areas of specialization, in their account of how information processing occurs.

The so-called "dual-coding" explanation of information processing continues to be explored in the context of learning. Mayer (2010), for example, noted the benefit of applying the basic organizational principle through multimedia (pictures and words) in the instruction of medical students. One of Mayer's main goals was to reduce "extraneous" (unnecessary to learning) processing while enhancing "essential" (relevant) processing. In a more recent study, Brown (2015) successfully used an information processing approach to assist pre-service teachers learn about the concept of social stratification. The idea of eliminating exposure to irrelevant content readily applies to an overhaul of a federal training system with an end of increased efficiency in mind.

Interestingly, recent research in the area of cognitive load (roughly, the *amount* of information one can handle at a given moment) indicates it is unclear whether information processing can be so discretely categorized as either a binary or unified system. Morrison, Burnham, and Morrison (2015) suggested that a more appropriate description of information processing requires a potentially complex hierarchical structure. In sum, there is yet no consensus as to the functional architecture of human cognitive framework.

The idea that active interaction with incoming information is an important component of learning is not a new concept. In Kolb's (1984) classic work on "experiential learning," attention focuses toward the transformative potential of learners' active experiences. Extending the general principles of experiential learning, Tomkins and Ulus (2016) note that a successful learning environment likely requires 'experiential' participation, from both students *and* instructors. In spite of the varied narratives of how mental processes manage incoming information, lessons from the cognitivists' theoretical lineage are useful to consider when designing updated mandatory training models for governmental use.

Lower- and Higher-Order Processing

The idea of separating classes of cognitive processes according to levels of difficulty (or, "involvement") is not new. In Bloom's taxonomic structure, cognitive operations fall into categories roughly based on procedural sophistication (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). *Knowledge*, for example, is representative of many lower-order processing tasks, covering simple recognition and memorization of mostly factual information. Alternatively, *application* operations require learners to engage in higher-order processes such as mapping complex conceptual information to specific problem areas in novel situations. At present, the majority of testing effect literature addresses lower-order concerns with relatively uncomplicated paradigms involving word pairs and lists. Thus, a contribution of the current effort is to evaluate test-enhanced instruction in the higher-order domains of *application* and *evaluation*. It is noted that an updated version of Bloom's taxonomy exists (Anderson & Krathwohl, 2001). However, the inclusion of a "*creativity*" stratum was not deemed applicable in the

context of mandatory federal training environments, so the original taxonomic structure was employed as a guideline.

Bramwell-Lalor and Rainford (2014) highlight the formidable challenge facing teachers of catalyzing higher-order thinking among students. Perhaps in tacit response to this longstanding dilemma, much of the available explanatory evidence regarding the nature of higher-order cognition stems (logically) from memory-related research. This is to be expected within the diverse contextual settings comprising education across the lifespan. Unsworth and Engle (2005) claim that working memory capacity is directly related to attentional control (via central executive processing), and that resultant increased attentional involvement during learning events promotes greater informational retention. The relationship between working memory and higher-order cognitive ability may be exceedingly nuanced, though. For example, it is possible that processing speed competes with limited cognitive resources involved with remembering information, with faster retrieval speeds driving enhanced accuracy (Unsworth, Redick, Heitz, Broadway & Engle, 2009).

It is also speculated that "gist processing," essentially deriving meaning from complex information, is a suitable index for gauging higher-order thinking (Vas, Spence & Chapman, 2015). The authors contend memory-related executive processes (including inhibition and switching) responsible for capturing and synthesizing the gist of long text passages are sufficient to serve as indicators of high-level cognitive operation. A substantial number of questions in the current investigation into the testing effect on mandatory training require similar high-level information manipulation.

Age-related declines in higher-order cognition are well-documented (Verhaegen & Salthouse, 1997). Interestingly, Darowski, Helder, Zacks, Hasher and Hambrick (2008) looked at the role of distraction as a potential mediator of poorer cognitive performance as humans grow older. Specifically, they tested adults (ranging in age from 18-87) on generating synonyms for target words, and found the ability to mitigate distractive influence was a protective factor against performance decrements traditionally associated with aging. This finding is particularly important in the sphere of mandatory federal training. Government employees include adults throughout the lifespan, who are subject to increasingly pervasive attentional intrusion from technological devices (such as phones and personal computers) which may be required for work.

The present research addresses two main areas. First, as detailed previously, a primary line of inquiry concerns whether adding content checks during teaching events is a useful strategy to encourage learning. Secondly, this project serves as an investigation into the effectiveness of varying instructional modalities on later knowledge retention. A considerably promising aspect of pursuing the combined set of questions is the prospect of flagging a potential avenue of U.S. federal funding that may be worth revising. Wasteful government spending is a problem that affects the citizenry at large, and is a matter of broad territory. Indeed, even in the lauded and generally esteemed arena of scientific exploration, the questionable channeling of revenue abounds (Miller, 2013). In a time of ever encroaching federal deficits, it is important to identify governmental operations that can be streamlined to save money and improve efficiency. Mandatory federal employee training is a costly endeavor in need of substantial reform (GAO, 2012). Simplifying the production of training modules, thereby lessening the associated

logistical expenses inherent in the federal training system, could save U.S. taxpayers a great deal of money.

Training modules can be both expensive and time-consuming to construct, involving the expenditure of a great deal of human and fiducial resources. Additionally, with video-based instruction there is often a need for periodic updating. This ensures a reflection of the "current-state" of relevant content, while helping to avoid projecting an "outdated" appearance. The former concern is potentially damaging in practical terms, while the latter issue could cause a disconnect between trainees and the material. Also, the period spent away from one's daily work is an important component to consider, as employees are (in a manner of speaking) "unproductive" during time spent passively receiving instruction. Thus, the need to improve the quality and proficiency with which federal employees are trained stands evident.

When training is compulsory, the potential exists for the imposition of unforeseen consequences on learners. Mythen and Gidman (2011) highlighted the possibility that mandatory training undermines the quality of learning. In the context of professional healthcare environments, the authors go further in the explanation that involuntary participation effectively induces in the learner a "subordinate role" which may ultimately interfere with a basic desire to learn. Similarly, Curado, Henriques, and Ribeiro, (2015) observed a decrease in willingness to transfer training-based knowledge in a group of workers at a large insurance company. Encouragingly, current research in the field of test-enhanced learning suggests adaptably specific instructional practices to promote efficient knowledge transfer.

Significant improvement in training practices is known to produce dramatic positive outcomes. Gaskell, Hinton, Page, Elvins and Malin (2016) instituted a program to improve the efficiency of mandatory training for new, incoming physicians. By streamlining the content delivery system through incorporating online instructional elements, they not only shortened the period needed for training completion (effectively reducing the sessions by approximately 19.5 hours), but elevated patient safety in that the doctors were able to spend the additional time becoming acquainted with their new working surroundings.

It is my conclusion that test-enhanced instruction should be tried with federal employee training programs in an effort to address two initial issues of concern. First, a streamlined instructional protocol incorporating intermittent testing may have advantages above purely "informative" types of training methods. If so, then the benefit of increased effectiveness in module construction should allow for more prudent allocative decisions regarding federal funding and related labor hours, predicated on the proposition of more efficient learning transfer. Second, the motivational decrement tied to mandatory training remains a problematic matter. It is worth investigating the extent to which test-enhanced instruction may protect against negatively-valenced attitude shifts commonly seen as reactions to the compulsory nature of training requirements. Possibly through a heightened level of engagement with information contained in the instructional modules, employees will serve less as "passive storehouses" of knowledge, and more as "active partners" in the acquisition and creation of enduring cognitive changes. Active involvement in learning episodes is believed to be a critical component during exposure to new material (LoPresto & Slater, 2016).

Major Aims

The purpose of this project manifests in two separable (but related) overarching goals. The first major aim of this study is to investigate the phenomenon of the testing effect. Simply, the idea is founded on the notion that embedding periodic, abbreviated content checks promotes a greater degree of learning than merely "studying" the material through rehearsal or the implementation of alternate memory-based strategies. To help answer the question of whether test-enhanced learning provides advantages in the context of mandatory federal training, the independent variable of *instructional style* is separated into two categories. The "pure-content" conditions contain the relevant instructional information only, in a manner emblematic of traditional federal training modules. No content checks occur in the pure content conditions. Conversely, the "content + testing" conditions precisely mirror the original material from the pure content conditions, with supplementation by brief, intermittent "quiz" questions integrated into the flow of the content presentation.

Second, the current study addresses the question of whether different content delivery methods impact later learning to the same extent. The independent variable of *presentation method* is partitioned into three distinct presentation categories: text-only, audio-only, and video with accompanying Powerpoint slides. Typical testing effect studies using dynamic stimuli employ video-only conditions (Chan, Thomas, & Bulevich, 2009; LaPaglia & Chan, 2012; Gordon & Thomas, 2014). The current study is unique in that the USERRA video contains several powerpoint slides, thus making it a hybrid video-text presentation. The content is consistent (verbatim) across categories. The audio-only conditions consist of the audio track from the video, with questions inserted at

regular intervals for the testing version. The text-only conditions consist of word-for-word transcriptions of the video content. Evidence indicates that distinct stimulus modalities may activate memory systems in different ways (Ogden and Jones, 2011). Thus, presentation method could feasibly impact learning outcomes. For example, Hannon (2012) notes higher-level reading comprehension processing is linked to working memory function.

In the enterprise of mandatory federal training, each of the respective presentation methods is unique in terms of production cost. Text-only presentations are the simplest and most cost-effective to produce, followed by audio-only modules. The most labor-intensive presentation method involves arranging content in the form of a video-based instructional module supplemented with informational slides. By comparing the effectiveness of knowledge transfer across the three content presentation schemes, I am able to secure initial evidence of whether one presentation method is logistically preferential above others. In light of elevated processing speeds, I predict that participants in the text-only conditions will complete the training faster their audio/video-based counterparts. Whether the text-only subjects will acquire a greater degree of learning is less clear, as the literature on self-paced instruction is historically somewhat mixed, relative to preference and performance (Ainsworth, 1979; Weng, 2015; Preusser, Bartels, & Norstrom, 2011; Semb, Glick, & Spencer, 1979). Currently, a substantial portion of mandatory federal employee training occurs through video-based presentation. It is hoped that this project can inform a restructuring of the training system resulting in more efficient (and cost effective) learning among U.S. federal employees.

Chapter 2: Literature Review

Test-Enhanced Learning (the "Testing Effect")

The preceding section provided an introductory discussion of the view that humans' cognitive structures are preferentially arranged to process information. Against that backdrop, the focus of this section is to offer a consideration of one category of methods designed to enable the *transfer* of knowledge efficiently from instructors to students. Known alternately as the "testing effect," or "test-enhanced learning," the practice of [essentially] embedding frequent assessments within the schedule of instruction as a tool to promote better learning is viewed as an effective educational strategy (Roediger & Karpicke, 2006a). At face value, having students/learners actively engaged with material during learning sessions through repeated content checks seems to align satisfactorily with the requirements of learners as outlined within the cognitivist information processing tradition (Ertmer & Newby, 1993). Noted by Larsen, Butler, and Roediger (2008), the testing effect was historically thought to induce greater learning either through the strengthening of associations by memory processing during retrieval, or (more simply) through repeated exposure to content-specific information; the authors note the former view is currently in favor. Experimental elements typical of testing effect investigations are briefly outlined in Table 2.1.

Table 2.1: Typical Experimental Designs of Testing Effect Studies

Study	Population	Stimuli	Learning Phase	Test Delay
Arnold & McDermott (2013)	Undergraduates (U.S. university; *N*=173)	40 pictures (free recall)	~2 minutes	30s
Bornstein, Liebel & Scarberry (1998)	Undergraduates (U.S. university; *N*=111)	Single video (~1:30 min); detail recall	~1:30 minutes	Immediate; 5 minute delay
Bouwmeester & Verkoeijen (2011)	Elementary school students (Netherlands; age 7-13; *N*=131)	12 word lists (8 items each)	<30 minutes	Immediate; 1-week delay
Jaeger, Eisenkraemer & Stein (2015)	Elementary school students (Brazil; age 8-10; *N*=69)	Text passage (321 words)	Self-paced; 2-4 read-throughs of passage	Immediate; 1-week delay
Karpicke & Roediger (2007)	Undergraduates (U.S. university; *N*=48)	52 word pairs from GRE	<1 hour (not explicitly specified)	10 minutes; 2 days
Rogalski, Altmann & Rosenbek (2014)	Older adults (U.S.; age 60-75; *N*=48)	Expository text passages	~1 hour	Immediate; 24-hour delay
Rowland & DeLosh (2015)	Undergraduates (U.S. university; *N*=36)	6 word lists (20 items each)	<1 hour (not explicitly specified)	1-, 4-, or 8-minute delay
Tse, Balota & Roediger (2010)	Middle-older adults (U.S.; age 46-95; *N*=96)	16 pictures of faces (with name and occupation labels)	<1 hour (not explicitly specified)	1.5 hour delay

The effects of testing might not be limited to immediate recall following learning events. Roediger and Karpicke (2006b), for example, determined that content-relevant testing improved performance on memory tests across extended timespans (up to a week beyond the initial instructional session) relative to simply re-studying the same material. Test-enhanced learning also appears advantageous over other instructional methods. In a study comparing generative explanation techniques, where first-year medical students were directed to interact with presented material to create detailed explanations, against test-enhanced learning conditions, Larsen, Butler, and Roediger (2013) found that a test-enhanced strategy was associated more favorably with subsequent test performance. Other classroom-based research reveals that introductory psychology students are amenable to repeated testing throughout the learning process, and students for whom regular testing is *required* tend to outperform those students for whom quizzing is *optional* (Trumbo, Leiting, McDaniel, & Hodge, 2016).

The benefits of test-enhanced learning can be extended to online-learning environments as well. Wojcikowski and Kirk (2013) found significant improvement in subjects' ability to diagnose patients following the presentation of online content in a scenario reflecting testing-enhanced learning. The authors speculate that a potential mediating factor for the phenomenon may have been the inclusion of specific, immediate feedback following the tests. However, evidence indicates the effectiveness of test-enhanced learning can be witnessed apart from feedback. Thomas and McDaniel (2013) observed unique effects of testing and feedback when investigating "front-end control" processes. One such mechanism, "source-constrained retrieval," is believed to function by enabling thinkers to allocate greater focus on intended informational targets.

Reiterating the idea of effortful processing (Schneider & Shiffrin, 1977), Rowland's (2014) meta-analysis of testing effect literature supports a "bifurcated" model, where recalled items in testing scenarios are given preferential "strength" in memory over equivalent items which have merely been restudied. In sum, the testing effect seems sturdy and test-enhanced instruction warrants greater exploration as a learning tool in both traditional classroom and online settings. This characteristic flexibility suggests that methods augmented through test-enhancement might be adaptable for use with mandatory federal government training, which transpires both "in house" (face-to-face) and remotely (virtually).

Of note, the literature regarding the necessity of feedback in mediating the testing effect is mixed. Some researchers assert that providing test-takers with timely information relevant to their performance is crucial to promoting the testing effect and subsequent learning (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Marsh, Fazio, & Goswick, 2012; Damhuis, Segers, & Verhoeven, 2015). In contrast, many fruitful demonstrations of test-enhanced instruction employ paradigms that do not include feedback mechanisms (Arnold & McDermott, 2013; Barber, Rajaram, & Marsh, 2008; Zaromb & Roediger, 2010; Putnam & Roediger, 2013; Henkel, 2007). Due to time constraints, the USERRA questionnaire was administered without feedback.

Presentation Method

A major manipulated component in the present design is the variable of presentation method. Simply, one of the main questions framing the current study is whether an observable difference in learning depends more heavily on a particular method of content presentation: text-only, audio-only, or video + related slides. As

detailed subsequently, a cursory summary of the literature on presentation method offers a markedly kaleidoscopic picture of evidence. In keeping with the spirit of the field, as it were, this detail of the larger current inquiry is rather unique in that it simultaneously includes three distinct presentation modalities.

When attempting to disentangle the role of text-based content presentation on later learning, the running backdrop of context appears crucial. Investigating the impact of visual supplements in conjunction with textual information, Butcher (2006) tested participants on the heart and circulatory system. The inclusion of diagrams (either simple or complex, in this case) was associated with greater learning than was observed in the text-alone condition. Similarly, Limperos, Buckner, Kaufman, and Frisby (2015) provided learners with instructional materials related to the concept of optimal performance, or "flow" (Csikszenmihalyi, 1990) in text-only and text augmented with audio conditions. Participants demonstrated greater knowledge transfer, and reported higher positive affect measures, when audio was presented in conjunction with text. Not all work is conclusive as to the questionable effects of text-specific content presentation. Izmirli and Kurt (2016) studied the instruction of basic computer concepts using combined modalities of text/animation and narration/animation. No substantial learning differences were noted across conditions. While somewhat isolated, their findings suggest that the instructional contributions of text- vs. audio-based presentations function equivalently (at least when holding animation constant).

Generally, much of the extant literature supports incorporating audio-centric presentation modalities into instructional practice. Middleton (2016) advocates the use of audio "podcasts" (digital audio files made available for download on the internet),

analyzing a case where material was presented from the field of undergraduate computer science education. Middleton cites the creation of dynamic, "rich learning spaces" that foster greater connectivity between teachers and learners. The effects of audio-based information exchange may also be extendable to a "secondary" class of instruction. In the context of a learning environment centered on mathematics, Weld (2014) found that audio-based feedback on students' writing assignments helped to facilitate increased comprehension of mistakes and to promote corrective techniques throughout the course. A complete account of the literature notwithstanding, some evidence indicates the productive effects of audio-based presentation methods may be less clear. Interestingly, a review of research focusing on the usage of podcasts as instructional tools uncovered that while students seem to explicitly prefer getting information through podcasts, the question of whether they are effective in the encouragement of learning is far from settled (Hew, 2009).

Comparatively more dynamic than audio- or text-alone delivery methods, video-based presentation methods represent a particularly potent system for transmitting information effectively to wide audiences of learners/students. Video-based content presentation has been used in a wide variety of formats (e.g., educational television and distance education) for decades (Wetzel, Radtke & Stern, 1994). Evidence indicates that video-based presentation can enhance learning experiences compared against non-supplemented classroom instruction. Lancellotti, Thomas, and Kohli, (2016) collected data from a "principles of marketing" class, where students either attended a traditional version of the class, or went to a class with the benefit of access to supplemental material in the form of video files. On the outcome measure of the final course exam, students

who were in the video-supplement condition outperformed those in the classroom-only instruction group.

Importantly, presentation methods using video are demonstrably effective for higher-order processing tasks. Support for video-based content delivery also comes from the field of medical education. Jones, Doleman, and Lund (2013) found that students preferred video (as opposed to audio) presentation when viewing/listening to interactions between novice and expert physicians, citing convenience as a major desirable factor. Further, video-based presentation may promote an eagerness to learn complex material. Choi and Johnson (2005), for example, investigated whether content delivery through video was an effective strategy for training masters-level students to adopt research-based teaching techniques in the classroom. The authors observed improved learning outcomes, as well as increased reported motivational levels, when course content was offered in video form, as compared with a traditional text version.

The ability to apply knowledge taken from federal training sessions is a consistently emphasized long-term educational goal. Encouragingly, evidence indicates video-based presentation methods are assistive with skill-transfer. In a vocational-preparatory context, Donkor (2010) noted beneficial effects of video content presentation over print methods for Ghanese inmates learning masonry, and that students reported being highly satisfied with video presentation in a follow up study (Donkor, 2011). Learner satisfaction seems elemental for sustaining employee motivation during training events. And willingness and capability to implement newly acquired material on the job are key components for federal employees interested in maximizing their potential gains from continuing professional education.

In contrast, the perception of video-based presentation in other contexts is less optimistic. Wright, Shumway, Terry, and Bartholomew (2012) compared several different presentation methods (e.g., traditional lecture, collaborative, video-based) in a sample of junior high learners. Tasked with acquiring new software applications skills, groups of both students and teachers initially rated the likely effectiveness of video-based presentation methods as relatively poor, when held against putative success employing alternate methods. In partial alignment with low expectations at the outset, Wright et al. (2012) noted that video-based presentation was associated with students whose grades landed them somewhere near the middle of the pack on final scores.

It is worth reiterating that the focus on presentation method as part of this study is somewhat unique. A consequence of including three separate presentation methods in this branch of the investigation could be the potential lack of observed practical learning differences relative to the Uniformed Services Employment and Reemployment Rights Act (USERRA) training content. The ramifications of such a situation are substantial. Under such conditions, it is possible that mandatory federal training could undergo a cumulative, informed shift toward a cheaper, more efficient form of information production. This could drastically reduce spending and wisely steer the reallocation of federal resources to other, more indispensable areas of government operation.

Presentation Method Is Not Learning Style

Reviewing the aforementioned work from the subfields under the umbrella of content presentation methods, it may be reflexive to infer that the present design springs from a standpoint tacitly endorsing the principles behind the "learning styles" phenomenon; this is not the case. As a brief reminder, Dunn (1990) outlined much of the

basic premise of the learning styles concept, focusing on the need to adapt instructional techniques to the diversity of strengths (essentially, preference) of different "types" of students. Among her many claims was the idea that accommodating individuals' unique learning styles was linked with significant improvement in performance across academic (and presumably professional) settings.

While initially enticing to a culture which prides itself on individuation, the practice of structuring classrooms and curricula around learning styles has since lost considerable favor. In a comprehensive review, Pashler, McDaniel, Rohrer, and Bjork (2009) did observe a trend whereby learners expressed preferences for particular methods of content presentation. However, they note that none of the studies analyzed satisfied conditional propositions required for concluding that *performance* was actually dependent upon an individual's stated style preference, when compared against other methods of presentation. Further, the educational material industry is potentially lucrative and curiously regulated. The adoption of educational technology and teaching/learning models is essentially under state and local control, with very little involvement from the federal level ("The Federal Role," 2016). Pashler et al. (2009) go so far as to insinuate that commercial motives may be the primary driver of success assigned to packages (and sold to schools) tailored to individual learning styles, in lieu of strong theoretical and empirical support. In spite of ingrained preferences, the common belief currently is that effective knowledge transfer is likely to occur when the presentation method maps well to the type of content under study, such as verbal poetry instruction (Brown, Roediger, & McDaniel, 2014).

Exemplars

The research contexts discussed subsequently highlight the diversity with which the fundamental concepts of test-enhanced learning can be applied. As such, they represent an increasingly comprehensive tapestry of age groups and environments. The first collection is an investigation of the testing effect in K-12 populations. This portion of the literature is crucial as it encapsulates both the earlier stages of formal education, as well as later transitional phases where students are preparing for entrance into adult life. The second part evaluates the test-enhanced learning phenomenon in relation to traditional collegiate settings. This represents especially fertile intellectual ground as a substantial portion of research is conducted on university campuses and affiliated institutions. The final area of inquiry concerns the testing effect in adult (i.e., post-collegiate) learners. It is particularly useful to consider whether test-enhanced benefits are derived in practical and continuing-educational environments, as the implications are potentially far-reaching (including, but not limited to, economic and policy-oriented matters). In short, compelling evidence abounds in support of test-enhanced instruction.

Area I: K-12 Settings

The bulk of confirmatory test-enhanced learning research has involved college students. In an effort to expand the knowledge base, Lipowski, Pyc, Dunlosky and Rawson (2014) investigated whether the principle behind the testing effect could be observed with a much younger population. The authors recruited groups of 1st- and 3rd-graders from middle-class elementary schools in Northeast Ohio. Each group viewed images of familiar objects (e.g., teddy bear, grapes, sock) divided into four higher-level categories (e.g., body parts, zoo animals), and completed the list-learning activities in two

conditions: restudy and test-plus-restudy; experimenters provided feedback when children incorrectly identified items. Children were tested in both conditions on separate occasions spanning a week. In cued (by category) and un-cued recall tests, collapsed across grade levels, students demonstrated better recall following the test-plus-restudy trials when compared to scores taken after the isolated restudy trials. An obvious success of the project was to observe an instance of the testing effect in an understudied population. Interestingly, the researchers found it difficult to convince the group of first-graders of the benefits of testing. This suggests that resistance to the idea of testing as a technique to improve learning may emerge at a relatively early developmental period in formal learning environments.

Effects of test-enhanced learning occur cross-culturally during the K-12 period as well. Jaeger, Eisenkraemer, and Stein (2015) compared the effectiveness of testing versus "restudy" (more explicitly in this case, *rereading*) in a sample of 3rd-grade children from low- to middle-class families in southern Brazil. Students initially read an encyclopedic passage containing information related to the sun. After the initial exposure to the material, children either took part in a cued recall test, or simply re-read the passage two additional times. On related memory assessments administered a week later, children who participated in the cued testing condition substantially outperformed those who did not. The authors take these results to indicate that the extensive reach of the testing effect includes "complex, educationally-relevant" material.

In early elementary populations, cognitive-developmental processing limitations (necessarily) restrict the nature of inquiry; thus, much of the research is confined to simplistic recall tests. Rohrer, Taylor, and Sholar (2010) explored whether the testing

effect would persist if the "final" assessment given to students included novel (as opposed to identical instructional phase) content demands more characteristic of knowledge transfer scenarios. Fourth and 5th-graders from a private Floridian elementary school were tasked with labeling locations on two maps (linked to condition) during an introductory learning phase; all students took part in both the "test-study" and "study only" conditions. Performance on two final tests (one mimicking that during instruction, and another designed to elicit transfer) favored learning from the test-study condition beyond studying alone. The evidence points to a test-enhanced phenomenon that sufficiently covers *generative* displays of knowledge.

Dirkx, Kester, and Kirschner (2014) continued the shift from simple recall tests by examining test-enhanced learning in the context of applied procedures/principles in high school statistics. A group of Dutch high school students were shown content on probability calculations from an age-appropriate math textbook, and either studied the material only or engaged in periodic testing during the learning stage. At the conclusion of a one-week interval, subjects in the testing condition were better able to apply the content-relevant knowledge than those in the study only group. The authors draw meaning not only in that the testing effect was identified in a later-adolescent sample, but also that the nature of the final test (applying learned information to actually solve problems, apart from successful conceptual recognition) shows that test-enhanced instruction may facilitate greater depth of mental processing.

Area II: College-Level Studies

As with most branches of psychology involving learning, a great deal of existing research on the testing effect comes from the college-aged population. Baghdady,

Carnahan, Lam and Woods (2014) investigated test-enhanced instruction in a sample of undergraduate Canadian dental and dental hygiene students. Students viewed learning material, including radiographic images of "intrabony anomalies," and slides with accompanying audio recordings, and were placed in either a study or a test condition. All participants had to provide diagnoses for a series of (patient) radiograph images and recall specific intrabony anomaly features, both during learning and at a one-week follow-up. Students in the test-enhanced group performed better than students in the study-only condition on the diagnostic accuracy assessment across both time points. No practical differences were observed in scores on the intrabony anomaly recall test at either stage. Baghdady, et al. (2014) felt they were able to induce effortful retrieval processing through the test-enhanced condition. According to Roediger and Karpicke (2006a), the "retrieval hypothesis" explains the testing effect through active processing of memory-stored information. As detailed previously, effortful processing is thought to be a crucial functional component in the cognitivist approach.

Using psychology students from an American university, Agarwal, Karpicke, Kang, Roediger, and McDermott (2008) sought to address whether the *type* of test given during instruction influenced later learning to a noticeable degree. Participants initially studied textbook prose passages on a variety of topics (e.g., "arctic explorer," "fossils"), and were split into test versus restudy conditions. The testing group was further subdivided into those who either received open- or closed-book tests at the conclusion of the instructional portion of the study. Later testing revealed that in both iterations of the testing condition, students fared better than study only counterparts on subsequent recall tests at a later time point. Agarwal et al. (2008) promote the theory that increasingly

rigorous challenges can lead to better learning, through Bjork's (1994) concept of "desirable difficulties." Notably, as in the earlier example involving first graders (Lipowski, Pyc, Dunlosky & Rawson, 2014), students were largely endowed with the notion that restudying would prove to be a more effective method for inducing later retention. The authors speculate that this (seemingly developmentally-consistent) phenomenon is due to metacognitive misinterpretation on the part of learners as to which strategies are better suited for successfully increasing sustained knowledge transfer.

Returning specifically to the question of how to understand factors that encourage learning transfer, Son and Rivas (2016) addressed the matter using the common classroom electronic "i-clicker" tool. Two groups of American introductory psychology students all used i-clickers for regular testing as part of typically-scheduled semester-long instruction. Those in the test-enhanced condition received additional questions, while students in the "notes" condition were given the same extra material to study. On the final exam, participants who regularly took part in the additional testing scored higher than the students whose instruction was only augmented with content-matching study materials. Son and Rivas conclude that their findings offer further support that test-enhanced instruction is advantageous in situations involving the transfer of acquired knowledge.

The testing effect has also been evaluated using thematic variations of visual recognition. Coppens, Verkoeijen, and Rikers (2011) had undergraduate students from the Netherlands learn word-symbol pairs, where the symbols were "Adinkra" images taken from the Asante tribe in Ghana. Consistent with other investigations of the testing effect, students were separated into groups according to test and restudy conditions.

Immediately following instruction, there was no discernible difference in performance between participants across conditions. However, at a one-week retest event, those originally placed in the testing group significantly outperformed their study-only counterparts. The authors interpreted the evidence as speaking to the generalizable nature of test-enhanced learning, as an extension to symbol recognition had been theretofore untested.

Area III: Professional and Informal Environments

Test-enhanced learning scenarios are suitable for use both in professional and less-traditional learning arenas. In a study involving first-year anesthesia residents, Galvagno Jr. and Segal (2009) investigated whether a testing intervention would promote learning. Participants (first-year residents from the Brigham and Women's Hospital training program) completed critical action procedures tests (CAPs) during the initial meeting to capture a baseline measure, and again at 1, 2, 4, and 9-month intervals thereafter. The CAPs were designed to cover "critical and essential actions," which if performed incorrectly or misremembered in *actual* working conditions could result in patients' deaths. Notably, CAP score percentages improved significantly from the baseline assessment to the 9-month follow-up testing session. As the study was observational and lacked a control group, the comparisons reflect within-group measures only. Participants generally viewed the repeated testing to be a valuable instructional practice. As the regular use of CAP-type tests has typically been restricted to potentially precarious occupations such as high-performance U.S. aircraft pilots, the authors counted as a success the informative adaptation of the testing scheme to the anesthesia residency program.

Meyer and Logan (2013) provide valuable data from a cross-sectional study on test-enhanced learning across adulthood. Groups included college students, similarly aged non-students, and middle-older adults (age 55-65). Participants interacted with previously unlearned material (articles about armadillos and black holes, for example) in the acquisition phase. As part of a counterbalanced within-subjects manipulation, each person participated in both the restudy and testing conditions (half of the content was used for each condition, respectively). Each participant was either tested a final time shortly following the conclusion of the learning phase, or two days after the event. In both delay conditions, each group of people exhibited greater learning in reference to the material included as part of the testing component (compared with information restudied) during instruction. Similar results have been obtained across a 1-week follow-up; see Kubik, Nilsson, Olofsson & Jonsson (2015) for further details. Demonstrative of a potential ability to employ test-enhanced learning throughout the lifespan, Meyer and Logan touted the applicability to benefit "nontraditional" students as well as the potential compatibility of test-enhanced instruction with career-based skills training. The authors suggested that *relevance* of the material to one's life (in work *or* education, presumably) may be a conduit through which the testing effect is delivered, so to speak.

Some evidence indicates that the testing effect can be enhanced through the use of timely feedback on performance. Tse, Balota, and Roediger (2010) studied healthy middle- to older-adults (age 46-95), tasking them with learning novel face-name pair associations. Individuals participated in both the restudy and testing versions of the acquisition phase. In one experiment, subjects received performance-based feedback; in a second part of similar design, no feedback was provided. When feedback was given,

both middle- and older-adults demonstrated a greater amount of learning on the final test, administered shortly following the completion of the acquisition phase. Curiously, in the absence of feedback, middle-adults displayed advantages in the testing condition, while older-adults displayed relatively better performance following restudy. At face value this may be construed as an age-range limited effect of test-enhanced instruction. To clarify, the authors argued that feedback may be an important component to consider when trying to identify the effects of testing in older populations, as they are more likely to possess fewer available cognitive resources.

An alternate perspective from which to gauge learning is found when addressing the issue in the opposite (degradative) direction. For example, Wheeler (2000) compared the rate of *forgetting* in younger and older adults. Participants were shown lists of categorical words; half of the individuals took part in a recall test following acquisition, while the other half of the sample did not. Expectedly, it was found that the rate of forgetting in older adults exceeded that of the younger adults when tested at after a 1-hour interval. Promisingly, individuals in the testing condition evidenced greater recall than non-test peers; this effect occurred across both age groups. Importantly, this could be viewed as test-enhanced learning invoking a protective influence against cognitive decline. Taken together, the collective body of work outlined above demonstrates the relevance of the testing effect to a diverse mix of groups over a wide range of contexts.

Chapter 3: Method

Participants

A total of 383 English-speaking adults took part in the study. This represents a response rate of 16.9%, as 2,266 individual recruitment emails were sent to solicit participation. Due to federal government guidelines, no information regarding age, gender, or race was collected. Participants were drawn from a population of federal legal employees, including: criminal attorneys, civil attorneys, professional support staff (e.g., paralegal specialists) and administrative support staff (e.g., budget officers, human resources personnel). In addition to occupational designations, participants provided information regarding the length of time in their current positions (e.g., less than 1 year, 10 or more years). Recruitment for the study occurred through email, and I personally handled all correspondence and scheduling matters. Participation was voluntary, and no compensation was provided in exchange for completing the study.

Materials

The content used in the current study comes from an existing federal legal training module. Specifically, the module covers the legal rules and regulations implemented through the passage of the Uniformed Services Employment and Reemployment Rights Act of 1994 (USERRA). The USERRA module is a mandatory training requirement for certain supervisory federal personnel. Its content describes in detail the rights and responsibilities of both employees and employers with regard to

handling situations in which current (or prospective) employees may experience work disruptions due to temporary military deployment. Essentially, the act protects the employment rights of members of the armed forces (e.g., full-time, Reservist, National Guard).

Some mandatory federal training is required of all employees on an annual basis. Uniquely, instruction on the USERRA Act is reserved for supervisory employees only, and thus does not occur on a repeating yearly schedule for all trainees. The aspect of exclusivity might protect against practice effects, contributing to a more authentic picture of the nature of learning in this scenario. Participants did not undergo a formal pre-test concerning USERRA content. However, they were asked whether they had previous exposure to the training module.

In its original form, the module is a video-based training course, with supplemental Powerpoint slides highlighting select key information. For the "audio-only" and "audio + test" conditions, the soundtrack to the module was used. This served to equivocate the video and audio conditions on the dimension of time length in the non-testing conditions. The content for the "text-only" and "text + test" conditions included verbatim transcriptions of the existing USERRA module, which I completed prior to data collection.

All questions, including those in the embedded content checks, were developed under the guidance of two subject matter experts. The USERRA video is 16 minutes and 47 seconds in length, and eight questions were interspersed throughout at roughly equal time points to create the "video + test" condition. Questions logically followed portions of the video, segmented in terms of both time and content. In the "audio + test"

condition, pauses were inserted into the audio track at the precise time points matching those in the "video + test" condition. For the "text + test" condition, breaks in the document directing participants to answer the content checks were matched to the "video + test" condition according to the transcript.

The embedded content checks consisted of eight total questions, varied according to different processing levels detailed in Bloom's taxonomic architecture (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). The questions were formatted as either multiple-choice or True/False. An example of a lower-order processing question is, "True or false: USERRA guarantees the right to reemployment for reservists and National Guard Members." This question represents the knowledge domain of Bloom's hierarchy, as it requires simple remembrance of factual detail. An example of a more demanding higher-order question is "After the receipt of notification of upcoming service, employers must do all of the following, except___." This question delves into an increasingly sophisticated set of premises, requiring participants to apply rules and regulations in a generic, hypothetical set of circumstances.

The full final assessment is composed of twenty-one questions. As with the embedded content checks, the set of final questions represents varying levels of cognitive demand in line with Bloom's taxonomic structure (Bloom et al., 1956). An example of a lower-order final question is, "the annual leave accrual rate includes the provision of up to ____ additional days for emergency service." This simply tasks the participant with recalling basic (though specific) factual information. Conversely, an example of a higher-order final assessment question is:

> "Jordan files a complaint with the Merit Systems Protection Board (MSPB) alleging discrimination against her employer. She was wounded while serving in

> Afghanistan, and convalesced for a period of 18 months. Upon attempting to return to her old job at the Department of Justice (DoJ), Jordan's supervisor informed her that the position had been permanently filled 6 months prior. Jordan's claim is not upheld, and she feels the situation was not handled properly. If she decides to appeal her case in District Court, should she reasonably expect her attorney(s)' fees to be covered under USERRA provisions?"

The latter example is a forced-choice dichotomous item that requires participants to simultaneously deal with both factual and applicatory styles of processing in order to successfully evaluate the fictional situation. The final assessment includes multiple-choice and True/False questions (some situational in nature), as well as an explanatory short-answer item. All multiple-choice, True/False, and situational questions were constructed to have a single correct answer, with no partially correct alternatives. I evaluated all answers to the single short-answer question in adherence to the USERRA guidelines. Part of the final assessment contained questions designed to assess participants' attitudes regarding the usefulness of the various presentation methods.

Following the content portion of the final assessment, three questions were included to gauge participants' attitudes toward the USERRA training module. Of most interest in the current study was the item: "The format of the training helped to facilitate my understanding of the material." Three demographic questions were added to the end of the assessment, asking participants to identify their general roles within offices/districts, the length of time they have been in those positions, and whether or not they had been exposed to the USERRA content prior to the study. Participants in all six conditions completed the final assessment, attitude and demographic sections. Those individuals in the three "testing" conditions completed the additional eight questions during the training segment of the study, before proceeding to the final assessment.

The USERRA video was displayed through a Sony VPL-FX30 projector onto a large screen at the front of a computer lab. Audio conditions were run through the A/V system in the same location. Both the set of interspersed content checks and the final assessments were presented through SNAP v.11, which is a software interface designed specifically to facilitate survey construction. The computer lab was outfitted with HP ProBook 640 G1 laptops.

Lower-Order vs. Higher-Order

Due to low item-total correlations, two questions (10 & 11) were removed from the final assessment prior to analyses. The entire assessment, including the in-presentation testing questions, is contained in Appendix A. In conjunction with Bloom's taxonomic structure, (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956), USERRA questionnaire items were parsed according to processing depth. The lower-order (LO) final assessment items included questions 9, 12, 13, 15, 17, 18, 19, 20, 21, 22, and 23. An example of a lower-order item was:

> "True or false: Title 10 status refers to National Guard members performing *state* duties."

This type of item is reflective of simplistic memorization operations, and challenges participants to recall specific informational guidelines. The higher-order (HO) items on the final assessment included questions 14, 16, 24, 25, 26, 27, 28, and 29. An example of a higher-order item is:

> "Rosita is an attorney with the DoJ and is currently deployed as a member of the National Guard in state service to provide hurricane relief to coastal communities on the gulf coast of Florida. Rosita learns from a superior that her deployment will end in five days. She calls her HR representative at the DoJ and verbally conveys her intent to return to work following the end of her deployment. Has Rosita provided sufficient notification to her employer under the USERRA guidelines?"

This kind of item requires participants to *apply* knowledge from the training module, while concurrently maintaining mental representations of multiple premises with potentially divergent outcomes.

Reliability

Prior to conducting the full-length study, the final assessment questions were piloted to a select group of participants (N = 15). A measure of internal consistency was calculated using Cronbach's alpha. The scale exhibited a moderate level (α=.56) of reliability using Cronbach's method. Additionally, reliability was checked with a split-half procedure, where the survey was divided into even and odd questions. The split-half reliability of the piloted assessment was moderate (r = .52), and the subsequent Spearman-Brown estimation of full reliability approached the high end of moderate (r = .68).

Reliability checks on the full-scale study (N = 383) yielded similar results, even after the removal of items 10 and 11. Overall, Cronbach's estimate of internal consistency for the final assessment (α=.55) was nearly identical to that observed in the pilot study. This supports the notion that the final USERRA survey carries a reasonably moderate level of internal consistency. Reliability of the full assessment was also checked using two separate split-half calculations. First, split-half reliability was determined for comparing the LO-HO questions. The correlation between question sets was moderate (r = .34), and returned a Spearman-Brown estimation of full-scale reliability of r = .51, which is within the moderate range. For comparison, an "even-odd" split-half reliability check was also performed. Under that condition, the correlation between forms was slightly lower than the HO-LO value, settling in the low range (r =

.27). Estimating the full-scale reliability of an even-odd evaluation, the Spearman-Brown formula returned a value of $r = .42$, demonstrating markedly lower reliability in comparison to the LO-HO split.

Individual reliability analyses were performed on the separate banks of HO and LO questions. For the lower-order questions, the internal consistency fell in the low range, with a Cronbach's value of $\alpha=.39$. Concerning higher-order questions, the measure of internal consistency was similarly low ($\alpha=.32$). This issue, along with the potential impact on validity, is addressed in the discussion section.

Experimental Design

Each of the two categorical instructional styles ("test" and "no-test") is mapped separately onto the three presentation methods ("text," "audio," and "video"). The resulting design allows for an examination of the testing effect across the varying presentation schemes. Overall, the design is experimental in nature, with random assignment and control conditions built in to address both major research aims. The single dependent variable common to all participants is the score on a USERRA-based test, administered immediately following completion of the instructional module. Contributing to the content validity associated with the assessment, *two* attorneys working for the United States Department of Justice (DOJ) provided helpful guidance and feedback as subject matter experts (SMEs) in the development of the questions for the USERRA assessment. Further, validation of the end-of-module test occurred in accordance with the principles and practices outlined in Dr. Robert Johnson's *Constructing Cognitive Instruments* (EDRM 721) course in the College of Education at the University of South Carolina.

The primary dependent measure of interest in the current project is manifested as mean scores on the USERRA assessment, common across all participants. As such, an appropriate statistical tool for evaluating the data was an analysis of variance (ANOVA). The design included the independent variables of *instructional style* and *presentation method*. Instructional style is separated into two levels ("test" and "no-test"), while presentation method is parsed into three distinct categories ("text," "audio", and "video"), creating a total of six conditions. Each participant was randomly assigned to undergo instruction and testing in a single condition. Therefore, the study's primary design for analytical purposes is a 2 x 3 between-subjects ANOVA framework. The structure is graphically displayed in Table 3.1, with conditions labeled "I-VI." Scores on the "content checks" are viewed as peripheral, and will not be part of the primary analysis. Internal consistency was checked using Cronbach's coefficient alpha, and applying the Spearman-Brown Prophecy formula provided a measure of split-half reliability. Subject matter experts (SMEs) were consulted during the development of all the USERRA test questions. The informed guidance of practicing professionals who constructed the USERRA training module was instrumental in helping to ensure/improve construct validity.

Table 3.1: 2 x 3 Between-Subjects ANOVA Design

	Text-Only	Audio-Only	Video w/Powerpoint
Pure Content	I	II	III
Content + Testing	IV	V	VI

Participants completed the learning and testing at the National Advocacy Center (NAC) and were not compensated for taking part in the study. Visiting United States

Attorneys and other associated legal professionals were the population of interest. Recruitment involved resident (in-house) professionals, as well as pulling from those expected to travel to the NAC for various specialized training events during the period of data collection (May through July of 2017). Due to federal government restrictions, limited demographic information was collected for each participant. Each subject was randomly assigned to a single experimental condition.

The USERRA training video is relatively brief in length. The abbreviated learning phase is consistent with prior demonstrations of the testing effect (Arnold & McDermott, 2013; Bornstein, Liebel, & Scarberry, 1998; Pastotter, Weber, & Bauml, 2013). Additionally, there is empirical support for administering tests immediately following learning events (Bouwmeester & Verkoeijen, 2011; Rogalski, Altmann, & Rosenbek, 2014; Rowland & DeLosh, 2015). The two audio conditions (test and no-test) were composed entirely of the speaking track from the training video to help ensure temporal equivalency across audio and visual presentation methods. That is, text conditions were not restricted to a timeline to match the audio and visual presentations. One of the implicit questions of this investigation is whether simply reading relevant material is advantageous (in terms of time cost) over receiving the same content in through other delivery systems. I feel that forcing the text conditions into temporal adherence with the audio and visual portions of the design would have been an artificial construction that might have ultimately harmed data integrity. The amount of time participants need to complete the text-only conditions was measured, as it offers valuable insight into the efficiency aspect of the learning process according to presentation method.

Power Analysis

In a 2012 review, Phelps noted that the range of typical effect sizes from the testing effect literature normally begins at approximately .55 on the low end, bracketed by .88 on the higher end. The current study represents an examination of the testing effect on a comparatively unique population in a novel environment. As such, a modest effect size estimation of .3 was chosen to uphold a conservative approach. Setting the alpha error probability at .05, and the likelihood of detecting an effect at .8, a power analysis was performed using G*Power v.3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007). In the 6-condition model given the aforementioned fixed parameters, a sample size of 90 participants (or, 15 per condition) would be necessary to ascertain a potential impact of testing as a contributing factor. Further, 111 participants would be needed to confidently assess whether presentation method was a significant factor. To check for an interaction between testing and presentation method, a sample size of 111 is recommended through G*Power. As 111 is not an even multiple of six, a minimum total sample of 114 was selected for the current project, distributed evenly as 19 per condition.

Procedure

The site of the study was a federal legal training facility located in Columbia, South Carolina. Throughout the year, continuing professional development classes are held at the facility. Participants travel from all parts of the country to receive specialized training in a number of specific legal areas. Classes are typically announced months in advance, and registration occurs several weeks before the training sessions begin. I utilized class roles for recruitment. Each attending member from courses running from May until late July of 2017 was contacted to solicit involvement in the study. The only

selective criteria imposed was that each potential subject was to be a Department of Justice (DOJ) employee. Prior to data collection, the University of South Carolina's Institutional Review Board (IRB) approved the study.

Experimental sessions were held between 3-5 days per week, depending on the length and overlap of courses (i.e., some classes were Monday through Wednesday, others Monday through Friday, etc.). The six experimental conditions were randomly assigned to the available slots, such that none was repeated until all six had been run in a full random sequence. Additionally, participants were randomly assigned to the daily slots, based on their overall availability tied to the duration of their course.

Each experimental session began promptly at 7:30AM in a private computer lab in the basement of the facility. Participants were mostly run in groups, though occasionally individuals completed the study in isolation. Due to the nature of class size, attendance, and willingness to participate, experimental sessions varied in size considerably one day to the next (min = 1, max = 24). Participants arrived, and were informed as to the general nature of the study. They were told the project was related to improving mandatory federal training. Participants were also instructed as to the logistical details of the condition in which they were taking part. For example, subjects in the "text + test" condition were told to read through the transcript (all relevant documents were preloaded on each computer before each session), switch to the SNAP v.11 survey at the cued time points in the document, and to complete the final assessment. Participants in the "video no-test" condition were simply instructed to view the video on the projection screen, and to fill out the final survey promptly after the video ended. In the "audio-" and "video + test" conditions, the media presentations were

paused while participants answered the interspersed questions, and resumed once all participants had provided answers.

To record the time taken for participants to read through the transcript, I moved to an observation position at the back of the computer lab. Immediately after giving the final instructions to begin the experiment, I started a stopwatch. During each session, I sketched out a participant seating chart, and recorded the amount of time for each person between starting the transcript reading and opening the final assessment. As the question concerned the potential advantage for the text-based presentation method, times were collapsed across the "test" and "no-test" conditions.

Most training classes start at 8:30AM. The study began each morning at 7:30, and ran between 30-40 minutes. Further, due to the self-paced nature of the final assessment, participants did not finish the study at similar times. In light of these two factors, no official debriefing occurred. Some participants stayed behind and wanted to discuss the purpose of the study in greater detail. During such occasions, I explained the concepts of the testing effect, as well as the rationale behind evaluating distinct presentation methods. Participants were asked not to share procedural or conceptual information from the study. Upon completion of the final assessment, participants were thanked for taking part in the study.

Analyses

Total scores on the complete set of items were analyzed using a univariate analysis of variance (ANOVA), with presentation method and instructional style included as independent measures. Additionally, the total bank of questions was split into lower- and higher-order processing items. Separate ANOVAs were performed with scores on

lower- and higher-order items as dependent variables, including presentation method, instructional style, and previous exposure to the USERRA content as independent variables. Finally, an analysis of variance was conducted on the attitudinal measure of whether the *format* of the training was believed to facilitate understanding of the material. Answers on the attitudinal item served as the dependent measure, while presentation method and instructional style were the independent variables.

Chapter 4: Results

Main Effects

Testing Effect

It was predicted that an overall effect of testing would be observed across conditions. A 3 x 2 analysis of variance (ANOVA) was conducted with overall scores on the USERRA assessment as the dependent measure, and presentation method (text, audio, video) and instructional style (test, no-test) as independent factors. As the model-level *F*-test did not approach significance, those findings are not reported. An interpretation of the absence of an effect of testing is found in the discussion section. Detailed analyses regarding the specific lower-level impacts of testing are outlined subsequently.

Presentation Method

Participants completed the study through one of three possible presentation methods: video-, audio-, or text-based modalities. It was expected that participants in the text-based conditions (both "test" and "no-test" versions) would finish the study in less time than their counterparts in the video and audio conditions. Generally, the average adult reads approximately 184 +/- 29 words per minute (Trauzettel-Klosinski & Dietz, 2012), while most people speak at a rate of around 150 words per minute (Reynolds & Givens, 2001). In accordance with these standards, participants in the current study completed the USERRA training module substantially faster when the material was presented in text format.

Overall, participants in the text conditions (n = 124) averaged 12 minutes and 19 seconds to read through the transcript. Recalling that the total time of the USERRA video (and by default, audio) is 16 minutes and 47 seconds, participants in the text condition completed the training module 27% faster across both text conditions than in the fixed-length video and audio versions.

Additionally, it was expected that differences between the overall learning outcome would not depend on presentation method. A 3 x 2 analysis of variance (ANOVA) was conducted with overall scores on the USERRA assessment as the dependent measure, and presentation method (text, audio, video) and instructional style (test, no-test) as independent factors. An interpretation of the absence of an effect of presentation method is found in detail the discussion section. Analyses addressing the specific interaction of presentation method and testing are outlined subsequently.

Lower-Order Findings

A three-way analysis of variance (ANOVA) was conducted with scores on the lower-order set of questions as the outcome variable, presentation method (text, audio, video), instructional style (test, no-test), and previous exposure to the USERRA content (no, yes) as the input variables. Presentation method was found to be a significant indicator of later performance on the lower-order portion of the USERRA assessment, $F(2, 371) = 6.46$, $p = .002$, $\eta^2 = .034$. Chiefly, participants exhibited the most favorable scores on the lower-order set of items in the video presentation method condition (M = .85, SD = .13), averaged across both instructional style levels. Specific pairwise comparisons indicated that subjects who viewed the USERRA video were significantly better at completing the final assessment than participants who either read the transcript

of (M = .78, SD = .14), or listened to the audio track (M = .81, SD = .14) of the same content. The overall difference in performance between participants in the audio and video conditions did not approach statistical significance.

To reiterate, no model-level effect of testing was observed for the current study for the overall assessment. Further analysis, however, revealed an interaction in performance between instructional style and presentation method, $F(2, 371) = 4.33$, p = .014, η^2 = .023. Particularly, when subjects were tested on the USERRA content as part of the instructional process, they tended to perform best on the lower-order bank of questions when the material was communicated through video presentation (M = .86, SD = .12). Participants who were tested while viewing the USERRA video significantly outperformed counterparts who were tested as part of the text-based (M = .75, SD = .14; p = .006) presentation format. Similarly, subjects in the "video + test" condition scored significantly better on the lower-order set of questions than other participants in the "audio + test" category (M = .81, SD = .14; p = .018). There was no significant distinction between performance on the lower-order questions for participants in the "text + test" and "audio + test" conditions.

For the lower-order set of questions, the analysis of variance revealed a significant effect of whether or not participants had previously been exposed to the USERRA training content, $F(1, 371) = 5.77$, p = .017, η^2 = .015. In concert with the findings for higher-order questions, participants who had seen the USERRA content before attending the experiment (M = .85, SD = .13) scored significantly higher than test-takers who were exposed to the USERRA content for the first time (M = .81, SD = .14).

Higher-Order Findings

As mentioned previously, model-level *F*-tests did not reveal significant effects of either presentation method or instructional style ("test" or "no-test"). Probing further, a three-way (3 x 2 x 2) analysis of variance (ANOVA) was performed on the set of higher-order questions, with presentation method, instructional style, and whether or not participants had previously seen the content (no, yes) as independent variables. Results revealed presentation method as a significant predictor of performance on higher-order items, $F(2, 371) = 4.75$, $p = .009$, $\eta^2 = .025$. Specifically, participants demonstrated the greatest amount of knowledge retention (on average) on higher-order questions when they completed the module in text-form ($M = .80$, $SD = .16$). Pairwise comparisons showed performance on higher-order items was significantly better ($p = .007$) for text-presentation over the audio format ($M = .75$, $SD = .18$). Performance differences between text and video ($M = .77$, $SD = .16$) formats, and audio and video formats were not significant.

While no main effect of testing emerged, there were differences within specific levels of the instructional style independent variable. For higher-order questions, testing was a significant predictor of performance in the context of presentation method, $F(2, 371) = 5.35$, $p = .005$, $\eta^2 = .028$. Participants scored best on higher-order items when tested as part of the text condition ($M = .82$, $SD = .17$). Pairwise comparisons indicate testing with text was significantly more effective at promoting retention than either testing with audio ($M = .75$, $SD = .18$; $p = .001$), or testing with video ($M = .77$, $SD = .16$; $p = .036$). There was no significant score difference observed between the test + audio and test + video conditions.

Part of the analysis of variance included a test to discern any potential effects of having previously seen the USERRA training content. Previous exposure to the training module produced a significant effect on performance, $F(1, 371) = 7.49$, $p = .007$, $\eta^2 = .02$. Participants who had previously been exposed to the USERRA content ($M = .81$, $SD = .15$) substantially outperformed their counterparts who had no prior knowledge of the USERRA content ($M = .76$, $SD = .17$).

Perceived Effectiveness

To address whether presentation method impacted the perceived effectiveness of the training event, an 3 x 2 analysis of variance (ANOVA) was conducted with the item as the outcome, and presentation method and instructional style as independent variables.

For overall scores on the question of whether the training format facilitated participants' understanding of the content, there was a main effect of delivery method, $F(2, 377) = 12.45$, $p < .001$, $\eta^2 = .062$. Participants favored the video-based presentation method ($M = 2.5$, $SE = .7$) primarily as a significant facilitator for learning the USERRA material. Pairwise tests highlight that video was rated significantly higher than both text ($M = 2.17$, $SD = .86$; $p = .002$) and audio ($M = 2.03$, $SD = .76$; $p < .001$) presentation formats. There was no significant difference between participants' ratings of audio compared to text formats for the facilitation of understanding the material. There was no significant effect of instructional style ("test" vs. "no-test") on whether participants viewed either style as helping them learn the USERRA content.

On whether presentation method facilitated content understanding, a significant interaction was observed between presentation method and instructional style, $F(2, 377) = 4.61$, $p = .01$, $\eta^2 = .024$. Within the text conditions, participants who were tested ($M =$

2.3, *SD* = .89) felt more strongly that the instructional style facilitated their understanding of the material than their counterparts who did not undergo testing (*M* = 2.02, *SD* = .81). This trend was reversed in the audio conditions, where participants who listened to the audio track felt the "no-test" style (*M* = 2.15, *SD* = .7) was preferential for learning over being tested (*M* = 1.93, *SD* = .81). Similar to people in the audio conditions, participants who viewed the video-based USERRA training format reported a greater perception of learning facilitation when they were not tested (*M* = 2.62, *SD* = .67) as opposed to participants who were tested (*M* = 2.38, *SD* = .71).

Chapter 5: Discussion

Testing Effect

As noted previously, there was no main effect of instructional style in the current study. That is, participants in aggregate across levels of the presentation method variable did not perform significantly different as groups on the final USERRA assessment based on whether or not they were tested during the learning phase. Given that the testing effect has been observed in multiple settings, and on several points along the human developmental continuum, the lack of a main effect reflecting a positive impact of testing was somewhat unexpected. There are several reasons why the current investigation may have failed to produce an overall educational benefit from testing during instruction.

First, it was possible that the participants' educational backgrounds might have obscured a typical effect of test-enhanced instruction. Subjects in the current study were drawn from high-level legal workplaces. Some variability in the nature of occupational classification was present, as participants included both civil and criminal Assistant United States Attorneys (AUSAs), legal support staff (e.g., paralegals, budget specialists), and administrative support personnel (e.g., human resource representatives). An important commonality among the occupations is the requirement of extensive education, from law school to bachelor's and master's degrees.

The current investigation serves as a glimpse into a particularly unique educational environment. The bulk of the testing effect literature represents studies

conducted using college students. However, there is precedent that professionals may benefit from testing during instruction. Galvagno Jr. and Segal (2009) observed improved performance for anesthesiology residents on the outcome of "critical action patterns" (CAPs), lifesaving medical protocols, for residents who had undergone regular assessments since initially learning the crucial sets of steps. As such, participants may have exhibited a generalized "practice effect" of sorts. It could be the case that through extended periods of formal education, participants simply developed adaptive test-taking strategies. Groups in the current study represent actual hires from the selected occupations, so it is reasonable to assume that they achieved desirable levels of test performance along the way to degree completion.

Another potential explanation for the lack of a benefit from testing relates to content difficulty. The content for the USERRA assessment was developed in conjunction with feedback from two subject matter experts (SMEs), each of whom attested to the material's content validity. However, the two SMEs did not provide an assessment of the overall or specific item-related difficulty of the questions. Participants in both aggregate "test" and "no-test" conditions performed fairly well on the final assessment (78% and 80%, respectively). While the higher-order and lower-order segments of the test charged participants with marshaling presumably distinct sets of cognitive resources, it remains a possibility that the questions were not challenging enough to elicit a benefit of testing.

Finally, the delay between learning and testing may not have been sufficient to reveal a testing effect. Some evidence indicates that testing immediately after the learning phase is an appropriate way to gauge whether participants demonstrate increased

performance on assessments (Arnold & McDermott, 2013). Participants in the current study completed an abbreviated learning phase (~17 minutes), and were then given the final test with no delay following the learning event. The absence of an intervening period between learning and testing was an unavoidable logistical constraint in the present study. It is possible that implementing mandatory breaks of varying length could have successfully instigated pronounced memory-related differences in participants in the "test" versus "no-test" conditions.

For the lower-order set of questions, increased overall performance through testing was linked with a specific content presentation method. Participants in the "video + test" condition substantially outperformed their counterparts who were tested during either the audio or text content presentations. The lower-order questions on the final assessment are believed to require less extensive cognitive processing than the higher-order set of items. Due to the comparatively simplistic nature of the items, participants may have preferentially benefited from the richness of detail provided by completing the training with the original video module. The lower-order questions essentially required subjects to recall relevant facts and guidelines, limiting potential undue strain on working memory capacity. It is possible that participants who were tested during the traditional video instructional module gained an advantage through being exposed to the content in a comparatively dynamic fashion, relative to audio- and text-based subjects. The video modality might have reinforced learning through increased attentional capture, thus demonstrated through enhanced outcomes on the lower-order set of questions.

An interaction effect of testing and delivery method was observed when examining performance on higher-order items. Specifically, a difference emerged across

groups in the “test” conditions. When learning the content through a style that included testing as part of the regular instructional process, participants fared better when the training information was channeled through a text-based presentation method. This relationship occurred when compared to both audio- and video-based modalities. The finding of increased performance through text + testing might be indicative of participants’ protracted educational histories. Throughout most formal k-12 settings, testing typically occurs in the form of computerized (formerly written) assessments that rely heavily on textual prompts. Further, most course materials have traditionally consisted of a primary academic textbook. Therefore, participants might have grown accustomed to answering questions following the reading of large chunks of information throughout their educational development.

It may be the case that interspersing questions in video and audio presentations served to disrupt attentional processing, ultimately causing distraction instead of facilitating improvements in learning. There is evidence that interruptions in audio and video learning events can have effects on later retention, though the likelihood of predictable consistency is yet to be established. Shuyan, Kuschpel, Schad, Heinz and Rapp (2015), for example, studied word retention in audio and visual stimuli paradigms. They found that breaks during instruction (specifically, video games) harmed auditory-based learning effectiveness, while seemingly enhancing later retrieval from visually-oriented conditions. Notably, the breaks in the USERRA training were content-congruent, while those in the Shuyan et al. (2015) paper were unrelated to the learning task.

Interestingly, no significant performance differences were noted between "audio + test" and "video + test" conditions in the current study. This may indicate distinctions related to information encoding streams. Specifically, participants reading the material engaged in a self-paced, mainly visual manipulation of the information. Conversely, subjects in the "audio + test" and "video + test" conditions received the content through listening to the expert speakers featured in the original training module.

Presentation Method

There was no overall main effect of presentation method observed in the current study. That is, participants (collapsed across testing conditions) performed equally well whether they were exposed to the USERRA training content in either the text-, audio- or video-based presentation schemes. Of note, a specific interaction between modalities and instructional style was observed, discussed previously in greater detail.

As hypothesized at the outset, content presentation method did not seem an overly important factor in the current examination of mandatory federal training. This is encouraging, with at least one major wide-reaching implication. Generating comprehensive topical government training videos is an expensive and time-consuming practice. Production resources, such as outfitting an in-house broadcast studio with the necessary audio-visual equipment and requisite expert technical crew, can amount to significant financial cost. Additionally, for the sake of authenticity and informational veracity, subject matter experts (SMEs) are essential spokespeople and commentators for such videos. Their participation often involves travel to the centralized filming location, and results in time away from their primary set of DOJ responsibilities. Thus, ensuring the video training modules accurately portray the necessary important information for

trainees carries significant monetary and temporal costs. As the salaries of federal employees are funded through tax revenue, limiting the associated production expenses with creating mandatory training modules could amount to saving a great deal of taxpayers' money over time.

One must also consider the involvement of *trainees* in the continuing education effort at the federal level. Current regular practice involves thousands of government employees traveling to the centralized training facility yearly to undergo specialized legal instruction. As the bulk of attendees come from areas spread across the country, their involvement in the training process requires the federal government to shoulder the additional logistical costs of temporary accommodations related to rooming and dining concerns. The current study provides evidence of equivalence between three distinct content delivery methods: text, audio, and video. Extending this line of reasoning, transitioning more of the content into a text-based presentation format, rather than relying heavily on in-person instruction (as is typical of many training events), might prove an effective strategy for reducing the federal government's overall operating budget. It is reasonable to believe that such a rationale is applicable government-wide, and not strictly limited to the DOJ example outlined in this paper.

Beyond contributing to the discussion of fiduciary advantages, one still needs to ask *why* no significant differences in total performance were observed across distinct presentation methods. Multiple explanations for this finding are feasible. One possible reason for the lack of findings relates to content difficulty. It was mentioned previously that the content may have been too easy to result in significantly stratified scores according to the set of predictors. Without certain knowledge, however, one must allow

for the probability of the reverse being true. The USERRA module contains a substantial amount of very specific legal and procedural information, delivered in the traditional video format in a relatively abbreviated timespan (under 17 minutes). It could be the case that the level of informational manipulation incumbent upon participants was simply too great to evoke identifiable differences according to presentation type. The assessment questions might have been simply too challenging for performance differences to emerge at the aggregate level. In support of this idea, findings discussed previously suggest that difficulty level and presentation method might interact to drive learning outcomes in mandatory training exercises.

Another potential explanation for not finding a significant impact of presentation method concerns the notion that the three distinct methods are genuinely equally effective at promoting learning. Adults in various legal occupations at the federal level carry the experience of having navigated many years (sometimes decades) of dynamic formal learning environments, coupled with lifetimes of complex information processing and manipulation by virtue of being "thinking" beings. Further, the exponential technological growth over the last quarter century has resulted in a near constant inundation with information, albeit of questionable practical relevance in many circumstances. Nonetheless, a working adult in today's society cannot succeed without the pronounced ability to encode, recode and recall vast amounts of complex knowledge. That "perpetual practice" ingrained into professional and personal life may render the manner in which information is presented less crucial than it would have been in times before the ubiquitous proliferation of technological intrusion. Training designers and potential

instructors should take this possibility under advisement when constructing or updating training modules.

Previous Exposure

Some of the participants (n = 79) had viewed the USERRA training module prior to taking part in the current study as part of their regular work obligations. Data concerning the amount of time between having previously viewed the content and participating in the study were not collected. Such data would be an interesting contribution to a discussion on the nature of long-term memory decay. Overall, having previously viewed the USERRA module was linked with higher performance on the end-of-instruction assessment. This phenomenon was observed for both the lower-order and higher-order sets of items. The findings suggest that specific previous training potentially augments individuals' abilities to recall and manipulate statutory and procedural legal information, whether the tasks require basic lower-order "remembering" operations or deeper applications involving simultaneous awareness of the interrelatedness of multiple premises.

Perhaps it is expected to find that participants scored better on the final assessment if they had been exposed to the content at some point before completing the study. It is also reassuring for two potential reasons. First, on an individual level, it speaks to the attentiveness of DOJ employees in supervisory roles who are responsible for managing work-related issues for sets of subordinates (recall that the USERRA training module is only mandatory for those in supervisory roles). Had the employees not taken the initial training episodes seriously, it is speculative to assume they would have otherwise outscored participants for whom the material was truly novel. Second,

enhanced performance on the total outcome measure is a testament to the effectiveness of the original training module. If participants who had undergone USERRA training in years prior did not score better than their "uninitiated" counterparts in the present investigation, the efficacy of the original training format could have been called into question. Thus, department officials can appropriately retain a degree of confidence that the original version of the training module was successful at allowing students to learn, retain and apply complicated factual and propositional information.

Perceived Usefulness

Included in the attitude portion of the final assessment was an item that asked participants whether the format of the training experience facilitated their understanding of the material. Interestingly, subjective ratings indicated that the most preferable presentation method was video-based content delivery. Participants reported an overall significantly more favorable opinion of video presentation when compared to both text- and audio- delivery methods, collapsed across both testing conditions. Data collection regarding the age of subjects was prohibited as part of the approval process for the current study. It is tempting to consider that older individuals might not be as receptive to video-based instruction as their younger colleagues, having potentially completed much of their formal learning prior to the implementation of digital technology in the classroom. However, research suggests that older students might be exceedingly agreeable to the inclusion of video-based content presentation (Simonds & Brock, 2014).

Stimuli are thought to be more salient when presented in dynamic fashion, whether the information displayed represents threatening or non-threatening objects or events (Carretie et al., 2009). The content presented in the current study was decidedly

non-threatening. The content was heavily legal and procedural in nature. In addition, participants took part in the study at 7:30AM, having traveled to the training center from all over the country. Under such conditions, it is understandable that a presentation method employing dynamic visual stimuli would be more effective for capturing attention than less salient text- and audio-based methods.

Participants in this unique professional training environment clearly preferred to undergo learning by means of a video-oriented presentation method. This finding should be taken with a certain degree of caution. The observance of increased favorability ratings for video content presentation should not be considered as a mandate to convert all existing training to dynamic visual format, or to require that future modules employ video production efforts exclusively. As noted previously, no main effect of presentation method was seen in the current study. This aligns with the rationale distinguishing learning style and presentation method. To reiterate, just because individuals claim to benefit more in terms of understanding material delivered through one modality versus another, in no way does it guarantee they will ultimately demonstrate enhanced educational outcomes for having done so (Pashler, McDaniel, Rohrer, & Bjork, 2009).

Efficiency

One of the more practical conclusions to be drawn from the current effort relates to the temporal cost associated with mandatory training. The overall running time of the original USERRA training video is 16 minutes and 47 seconds. Participants in this study who completed the text-based transcript training presentation demonstrated a significant reduction in time over their counterparts who were exposed to the content through audio or video presentation methods. The effect was observed when collapsing across both

testing conditions. Overall, participants who read the transcript needed 12 minutes and 19 seconds to consume the material before beginning the final assessment. This resulted in a total time improvement of 4 minutes and 28 seconds above viewing (or listening to) the traditional USERRA module. Proportionally, the time it takes for readers to get through the material is around 27% less than the amount of time necessary to watch the original video.

At face value, this might seem relatively inconsequential. In the broader context, though, this finding is especially encouraging. The federal government employs over 2 ½ million civilian workers, according to a report by the Office of Personnel Management provided for the year 2013 (Office of Personnel Management, 2013). Expanded across that sizeable demographic, the potential for reducing the necessary time required to train federal personnel could have an additive effect on reducing government overspending. Employees in the current study appear faster (on average) at completing training when reading the content, and demonstrate no substantial comparative decrements in learning as a result. If this characterization accurately reflects a more holistic reality, then having federal employees undergo self-paced training in a text-based presentation method could conceivably allow for a greater portion of their regular time to be spent attending to the duties for which they were initially hired. When considering that federal employees must undergo regular annual training/retraining in a number of areas, the potential for increased efficiency through modifying the typical training format becomes ever more convincing.

Threats to Validity

The possibility of undue influence from external and internal confounding factors needs to be taken into account when interpreting the current results. It should be noted that the NAC is a centralized training facility for the Department of Justice. Therefore, attendees frequently cross time zones to fulfill continuing education requirements and receive basic/advanced skills instruction by traveling to the Columbia, SC, location. It is possible that jet lag, or other manifestations of general fatigue, affected participants' performance on the USERRA assessment. As part of a project to develop a scale designed to detect the occurrence of jet lag, Becker, Penzel and Fietze (2015) list "impairment of daytime functioning" as one of the primary screening criteria. Further, evidence suggests that the deleterious effects of jet lag may be akin to those experienced by workers on rotating shift schedules, including a reduced capacity on cognitive tasks involving vigilance and the maintenance of attentional processes (Akerstedt, 2007). It is reasonable to expect that the experience of jet lag-induced fatigue may have emerged in the form of either disrupted learning during the module training, or as a subsequent diminished set of stratified scores on the USERRA content test.

Additionally, various competing influences underlying motivation may have influenced the present findings. It is known that certain personality-based characteristics are (generally) thought to be displayed more commonly among those in legal professions, compared with the rest of the population. As Riech (2015) notes, lawyers are seemingly more prone to evidence behavior indicative of increased aggressiveness, competitiveness, and even psychopathic tendencies. Accordingly, these traits may contribute to a "general ambition to achieve," outstripping corresponding motivational drive in the larger

population. The potential for an elevated success drive needs to be taken into account when moving to apply the results from this study to more professionally heterogeneous groups of test takers/learners. It is hoped that randomizing to conditions worked toward minimizing unwanted interventional effects of motivation and personality characteristics.

Unlike the serious consequences inherent in poor performance outcomes associated with practical medical education (as in Galvagno Jr. & Segal, 2009), the current study involved no such potential for human harm. The present examination of a federal training module was a decidedly *low stakes* exercise. Additionally, in the absence of feedback, participants did not even have to deal with the possibility of acquiring knowledge of poor performance on the final USERRA assessment. In academic and professional settings, diminished ability can have real world effects, such as derailing progress toward degrees or termination from employment. With essentially "nothing to lose," participants may not have been sufficiently motivated to benefit from test-enhanced instruction.

It should be noted that while the training module used in the current study is mandatory for certain federal supervisory personnel, participation in the experiment was voluntary. As Rosenthal and Rosnow (1975) point out, people who volunteer for research studies tend to be more intelligent and highly educated than those who decline participation. Under that assumption, it is reasonable to believe that research volunteers may be more highly motivated than the general population. In the current context, this "self-selection bias" may have manifested as an equalizer of ability, where participants demonstrated equivalent performance undergirded by comparable levels of motivation.

The aspect of task difficulty should be kept at the forefront of mind when interpreting the outcome of the current research effort. The extent to which a learning session taxes one's cognitive resources is a significant determinant of achievement. Rice et al. (2012), for example, claimed that as task difficulty increases, performance is likely to suffer demonstrably by a degradation in consistency of observable skill and ability output. As noted previously, the USERRA training module is comparatively dense, and is heavily loaded with a litany of complex factual and procedural information. Thus, it is feasible that overloading learners with a more challenging task could have skewed scores to an extent, compared to other training modules requiring varying degrees of mental effort.

Finally, a major concern that could have negatively impacted the validity of the current findings relates to reliability. Cronbach's measure of internal consistency applied to the overall assessment produced a relatively low value (α=.55). It is difficult to claim with confidence that a survey scale is valid in the absence of strong reliability evidence. As related to the current effort, certain time constraints were in place that may have contributed to the suppression of internal consistency. The survey creation and pilot testing were carried out in accordance with Dr. Robert Johnson's *Constructing Cognitive Instruments* class in the Fall of 2016. As such, the development of scale and evaluative components had to follow a strict, set schedule. Further, the entirety of the project from conception to conclusion spanned approximately one year. Local administrators, as well as those in Washington, D.C., were eager to learn the results of the project. I felt it incumbent to deliver the findings in a reasonably timely manner. Under different, less time-sensitive circumstances, I would have conducted more extensive piloting of the final

assessment. The fact that the full-scale and pilot measure of internal consistency are equivalent speaks to the potential benefit of survey revision as a tool to improve reliability.

It is unclear how the present study compares to other investigations of test-enhanced instruction in terms of reliability. Apparently it is not common practice to report internal consistency metrics in the testing effect literature, and no estimation regarding an aggregate-level reliability of survey instruments is provided in Phelps' (2012) review. Nonetheless, it is noteworthy that the full-scale final USERRA assessment demonstrated a low level of internal consistency. This consideration may offer insight as to why the expected effect of test-augmented instruction was not identified in the current study.

An awareness of multidirectional threats to validity presumably operated to help reduce the likelihood that intervening variables interfered with, and undermined the authenticity of, the present study. While great care was taken to properly specify (i.e., model) expected relationships between important constructs, no study is ever truly impervious to the occurrence of unwanted effects caused by unanticipated factors or events. It is believed that the above brief admission of some of the more readily identifiable lurking "gremlins" ultimately contributed to a more comprehensive discussion of the vulnerability of the current design, and served to strengthen validity in the current study.

Limitations

A cursory review of the current project reveals a few limitations. First, it is important to recognize that the USERRA training module (and accompanying

assessment) represents a single content package. As such, the justification for inferential prowess is thus necessarily limited. The U.S. Department of Justice operates a training system that covers an extensively broad range of topics. It is possible that the pattern of correlations across conditions observed in relation to the USERRA test will not be consistent during the evaluation of additional learning modules. If so, the overall picture of *which* content presentation method is appropriate for facilitating knowledge transfer across disparate topics may be one of marked variability.

Material difficulty-level is another relevant consideration. Though the USERRA content module is relatively brief in duration (~17 minutes), it is densely packed with complex legal and procedural information. Comparatively, several of the other mandatory training units (e.g., sexual harassment, information security) represent knowledge areas which are routinely referred to anecdotally by federal employees as indicative of "common sense." This implies that they are generally geared for more simplistic knowledge acquisition. The resultant impression is that a substantial degree of variability exists across the collection of training modules with regard to the depth of information processing required to comprehend the various material. Applying the current research design to modules of differential challenge may reveal that particular content presentation methods are more amenable to specific, corresponding levels of difficulty.

The mandatory nature of the USERRA training module may also play into the mechanics underlying the testing effect. It is possible that learning outcomes in mandated scenarios will differ from results when employees have the option of participating in training sessions. Differences in performance are known to be linked

with varying kinds of motivation, which understandably may vary according to whether or not students volunteer for training. Further, intrinsic satisfaction is thought to be a significant driver of performance quality (Cerasoli, Nicklin, & Ford, 2014).

In terms of the ability to generalize to widespread learning contexts, this study is limited in that the sample will consist exclusively of adult professionals. The data were collected at the U.S. Department of Justice's National Advocacy Center. That confined responding to Assistant United States Attorneys (AUSAs) and their related legal and professional staff members. The highly field-specific nature of the training and education required to enter and maintain a career in the federal legal system initially distinguishes those who take part in the study from most of the rest of the U.S. population. Thus, interpreting conclusions from data produced in this study as applicable to additional educational and professional settings would be a precarious endeavor at best.

Admittedly, as is the case with most research ventures, the ability to translate findings to other contexts in a fruitful way will likely require thoughtful, nuanced adjustments to both protocol and content. In order to allow for proper comparisons of the testing effect in other populations, it is necessary to carry out additional work in areas further removed from the (relatively) compartmentalized pockets of government employees. For example, designs centered on more traditionally formal educational settings, as well as more practically-oriented environments (such as industrial, skills-based arenas), would provide elemental information to assist with evaluating the potential ubiquity of test-enhanced instruction.

Recommendation

Participants did not seem to benefit on the final USERRA assessment overall from testing versus not being tested on the material during instruction. Similarly, there was no significant distinction in performance observed across conditions of presentation method at the aggregate level. A slightly different picture emerged, however, when examining the interactions. To briefly recap, for the lower-order set of items subjects who received the USERRA content through the video-based presentation method outperformed participants who were exposed to the material through either the text- or audio-based presentation formats. Participants in the test condition seem to have largely influenced the learning advantage evidenced by subjects who experienced the video content presentation method. Conversely, participants scored better on the higher-order bank of questions when the information was presented in a text-based method. Similar to the observation with lower-order items, this effect seemed heavily influenced by individuals in the "text + test" condition.

I caution against abandoning the notion of test-enhanced instruction as a useful teaching strategy following the lack of finding a testing main effect in this study. Instead, I suggest that a nuanced approach designed in consideration with the varying degrees of item difficulty for a given assessment is warranted. Based on the present findings, when teachers and instructors are assembling training materials, they should do an initial parsing of potential questions according to Bloom's taxonomic structure (or an equivalent alternative measure) to determine putative cognitive processing levels likely to be catalyzed by each item. In this way, instructors will be able to tailor a mixed-modality content presentation system that can optimize learners' successful acquisition of detailed

knowledge. Based on ample evidence taken from studies of k-12 and college samples, the test-enhanced instruction enjoys a nearly comprehensive reputation for effectiveness. When applied to exceedingly specialized, non-traditional professional learning environments, soliciting advantages from the inclusion of tests during training sessions might require flexibly adaptive test construction.

Future Directions

The current study examined the impact of testing as an instructional style, in the context of varying content presentation methods (text, audio, and video). Both the site and the experimental design represent unique contributions to the investigation of test-enhanced instructional practice. Access to a centralized legal training facility managed by the Department of Justice provides entrance into a highly-specialized world of adult learners. No other population of potential research participants shares the combination of required training and professional obligations as those who took part in the present study. Further, I am aware of no other study that employed a comparison of three distinct content presentation methods when investigating the testing effect. The value of the current study is clear, and I believe justifies continued inquiry into the nature of test-enhanced instruction.

There is a great opportunity to expand this line of research within the Department of Justice. Numerous training modules exist covering a kaleidoscopic array of topics, from general sexual harassment to niche areas such as civil litigation. Revisions to existing training sequences and the creation of entirely new content are continual. The experimental framework detailed previously is readily amenable to other content areas beyond the USERRA module. It would be relatively simple to apply the structure to

investigate whether test-enhanced instruction is "content-dependent," and observable following training events addressing different subject matter.

In adapting the current research design for future productive use, it will be necessary to address the issue of temporal spacing between instructional events and testing sessions. To date, much of the work on the testing effect is understandably restricted to formal learning settings. Traditional k-12 and collegiate environments are ideal for studying educational outcomes. However, due to attrition, geographic mobility and assorted other factors, maintaining a cohort of students as willing research participants for a significantly lengthy amount of time is a thorny challenge. For an enduring effect of testing to appear, longitudinal designs are crucial. This is less important for condensed educational objectives, such as learning the chronological order of 17th-century British ruling families for a unit test on world history. Rather, where long-term evidence of the advantages of test-enhanced instruction needs to be explored is in the context of professional environments. Adults in career positions must demonstrate consistent knowledge mastery within their fields of expertise in order to maintain gainful employment. In the current study, participants were tested immediately following the instruction phase. I am interested in modifying the time between learning and testing. For professional purposes, it would be useful to examine performance on desired learning goals up to several years past the initial training episode.

As mentioned previously, the majority of literature on the testing effect comes from studies using k-12 and college samples. One aspect of the present study's contribution to the overall body of knowledge is the inclusion of a sample of highly-trained federal government personnel. As indicated by the present findings, whether the

testing effect is observable in professional adult populations is less clear than at earlier developmental time points along an individual's lifetime educational continuum. I would like to adapt the current paradigm for use with multiple adult professional samples from a broader spectrum of occupations. Careers requiring certification or special licensure, such as nurses and long-haul truck drivers, would be fruitful for investigating the potential impact of test-enhanced instruction. Part of the utility of test-enhanced instructional practice is the flexibility with which it can be incorporated into seemingly any professional situation. In consideration for "high-stakes" professions (e.g., commercial airline pilots, emergency room physicians, structural architects) where competency breakdowns or lapses might imperil lives, researchers should look to simulated assessments of work protocol wherever reasonably possible.

Another option for future work in this area is the use of non-parametric analyses to check for effects of testing and presentation method. For example, in the current study categorical data were collected regarding participants' occupational statuses. The present sample was too imbalanced to permit traditional linear modeling through ANOVA-based analytical methods, due to the violation of fundamental assumptions that would have rendered such testing irreversibly flawed. However, alternative procedures such as logistic regression would provide valuable information under conditions where traditional linear testing is inadvisable.

Finally, investigators going forward should focus on application when conducting research on the testing effect. Many of the studies in academic settings have tasked participants with merely remembering specific bits of information across relatively short intervals. From the simpler word lists and pairs, to the more elaborate reading

comprehension tests, a great deal of the evidence of the testing effect comes from learners' abilities to extract information from some set of stimuli. Of greater interest to adult working populations is the capacity for transferring relevant knowledge into practical job-based performance. Skills transfer is also of substantial import to k-12 and collegiate populations. It is likely that most teachers who take the time to plan detailed lessons want their students to be able to use the information beyond merely regurgitating it for chapter tests or final exams. Rather, it would benefit educators to know if their instructional methodology truly translates to adaptive student performance.

References

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*(7), 861-876.

Ainsworth, L.L. (1979). Self-paced instruction: An innovation that failed. *Teaching of Psychology, 6*(1), 42-46.

Akerstedt, T. (2007). Altered sleep/wake patterns and mental performance. *Physiology & Behavior, 90*(2-3), 209-218.

Anderson, L.W., & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Andre, T. (1972). An information processing theory of learning and forgetting, presented at The American Educational Research conference, Chicago, April 1972.

Arnold, K.M., & McDermott, K.B. (2013). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review, 20*, 507-513.

Baddeley, A.D., & Hitch, G.J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493.

Baghdady, M., Carnahan, H., Lam, E.W., & Woods, N.N. (2014). Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Medical Education, 48*, 181-188.

Bangert-Drowns, R.L., Kulik, C.C., Kulik, J.A., & Morgan, M. (1991). The instructional

effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Barber, S.J., Rajaram, S., & Marsh, E.J. (2008). Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience. *Memory, 16*(8), 934-946.

Becker, T., Penzel, T., & Fietze, I. (2015). A new German Charite jet lag scale for jet lag symptoms and application. *Ergonomics, 58*(5), 811-821.

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., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain.* New York, NY: Longmans, Green.

Bornstein, B.H., Liebel, L.M., & Scarberry, N.C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12*, 119-131.

Bouwmeester, S., & Verkoeijen, P.P.J.L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. *Journal of Memory and Language, 65*, 32-41.

Bramwell-Lalor, S., & Rainford, M. (2014). The effects of using concept mapping for improving advanced level biology students' lower- and higher-order cognitive skills. *International Journal of Science Education, 36*(5), 839-864.

Brown, J. (2015). Using information processing theory to teach social stratification to

pre-service teachers. *Journal of Education and Learning, 4*(4), 19-24.

Brown, P.C., Roediger, H.L., & McDaniel, M.A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: The Belknap Press of Harvard University Press.

Butcher, K.R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*(1), 182-197.

Carretie, L, Hinojosa, J.A., Lopez-Martin, S., Albert, J., Tapia, M., & Pozo, M.A. (2009). Danger is worse when it moves: Neural and behavioral indices of enhanced attentional capture by dynamic threatening stimuli. *Neuropsychologica, 47*(2), 364-369.

Cerasoli, C.P., Nicklin, J.M., Ford, M.T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980-1008.

Chan, J.C.K., Thomas, A.K., & Bulevich, J.B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20*(1), 66-73.

Choi, H.J., & Johnson, S.D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *American Journal of Distance Education, 19*(4), 215-227.

Coppens, L.C., Verkoeijen, P.P.J.L., & Rikers, R.M.J.P. (2011). Learning Adinkra symbols: the Effect of testing. *Journal of Cognitive Psychology, 23*(3), 351-357.

Csikszenmihalyi, M. (1990). *Flow: The psychology of optimal experience.* New York,

NY: Harper & Row.

Curado, C., Henriques, P.L., & Ribeiro, S. (2015). Voluntary or mandatory enrollment in training and the motivation to transfer learning. *International Journal of Training and Development, 19*(2), 98-109.

Damhuis, C.M.P., Segers, E., & Verhoeven, L. (2015). Stimulating breadth and depth of vocabulary via repeated storybook readings or tests. *School Effectiveness and School Improvement, 26*(3), 382-396.

Darowski, E.S., Helder, E., Zacks, R.T., Hasher, L., & Hambrick, D.Z. (2008). Age-related differences in cognition: The role of distraction control. *Neuropsychology, 22*(5), 638-644.

Dirkx, K.J.H., Kester, L., & Kirschner, P.A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research, 107*(5), 357-364.

Donkor, F. (2011). Assessment of learner acceptance and satisfaction with video-based instructional materials for teaching practical skills at a distance. *International Review of Research in Open and Distance Learning, 12*(5), 74-92.

Donkor, F. (2010). The comparative instructional effectiveness of print-based and video-based instructional materials for teaching practical skills at a distance. *International Review of Research in Open and Distance Learning, 11*(1), 96-116.

Dunn, R. (1990). Rita Dunn answers questions on learning styles. *Educational Leadership 48*(2), 15-19.

Ertmer, P.A., & Newby, T.J. (1993). Behaviorism, cognitivism, constructivism:

Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 6*(4), 50-72.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistic power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Galvagno Jr., S.M., & Segal, B.S. (2009). Critical action procedures testing: A novel method for test-enhanced learning. *Medical Education, 43*, 1182-1187.

Garraway, J., Volbrecht, T., Wicht, M., & Ximba, B. (2011). Transfer of knowledge between university and work. *Teaching in Higher Education, 16*(5), 529-540.

Gaskell, N., Hinton, R., Page, T., Elvins, T., & Malin, A. (2016). Putting an end to black wednesday: Improving patient safety by achieving comprehensive trust induction and mandatory training by day 1. *Clinical Medicine*, *16*(2), 124-128.

Gordon, L.T., & Thomas, A.K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory and Cognition, 42*, 186-197.

Hannon, B. (2012). Understanding the relative contributions of lower-level word processes, higher-level processes, and working memory to reading comprehension performance in proficient adult readers. *Reading Research Quarterly*, *47*(2), 125-152.

Heinrich, C.J. (2008). Advancing public sector performance analysis. *Applied Stochastic Models in Business and Government, 24*, 373-389.

Henkel, L.A. (2007). The benefits and costs of repeated memory tests for young and older adults. *Psychology and Aging, 22*(5), 580-595.

Hew, K.F. (2009). Use of audio podcast in K-12 and higher education: A review of

research topics and methodologies. *Educational Technology Research and Development, 57*(3), 333-357.

Hung, C.Y., Sun, J.C.Y., & Yu, P.T. (2015). The benefits of a challenge: Student motivation and flow experience in tablet-pc-game-based learning. *Interactive Learning Environments, 23*(2), 172-190.

Izmirli, S., & Kurt, A.A. (2016). Effects of modality and pace on achievement, mental effort, and positive affect in multimedia learning environments. *Journal of Educational Computing Research, 54*(3), 299-325.

Jaeger, A., Eisenkraemer, R.E., & Stein, L.M. (2015). Test-enhanced learning in third-grade children. *Educational Psychology, 35*(4), 513-521.

Jones, K., Doleman, B., & Lund, J. (2013). Dialogue vodcasts: A qualitative assessment. *Medical Education, 47*(11), 1130-1131.

Karpicke, J.D., & Roediger, H.L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704-719.

Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice Hall.

Kubik, V., Nilsson, L.G., Olofsson, J.K., & Jonsson, F.U. (2015). Effects of testing on subsequent re-encoding and long-term forgetting of action-relevant materials: On the influence of recall type. *Scandinavian Journal of Psychology, 56*, 475-481.

Lancellotti, M., Thomas, S., & Kohli, C. (2016). Online video modules for improvement in student learning. *Journal of Education for Business, 91*(1), 19-22.

LaPaglia, J.A., & Chan, J.C.K. (2013). Testing increases suggestibility for narrative-

based misinformation but reduces suggestibility for question-based misinformation. *Behavioral Sciences and the Law, 31,* 593-606.

Larsen, D.P., Butler, A.C., & Roediger, H.L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47,* 674-682.

Larsen, D.P., Butler, A.C., & Roediger, H.L. (2008). Test-enhanced learning in medical education. *Medical Education, 42,* 959-966.

Limperos, A.M., Buckner, M.M., Kaufman, R., & Frisby, B.N. (2015). Online teaching and technological affordances: An experimental investigation into the impact of modality and clarity on perceived and actual learning. *Computers & Education, 83,* 1-9.

Lipowski, S.L., Pyc, M.A., Dunlosky, J., & Rawson, K.A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology, 50*(4), 994-1000.

LoPresto, M.C., & Slater, T.F. (2016). A new comparison of active learning strategies to traditional lectures for teaching college astronomy. *Journal of Astronomy & Earth Sciences Education, 3*(1), 59-76.

Marsh, E.J., Fazio, L.K., & Goswick, A.E. (2012). Memorial consequences of testing school-aged children. *Memory, 20*(8), 899-906.

Mayer, R.E. (2010). Applying the science of learning to medical education. *Medical Education, 44,* 543-549.

Meyer, A.N.D., & Logan, J.M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging, 28*(1), 142-147.

Middleton, A. (2016). Reconsidering the role of recorded audio as a rich, flexible and engaging learning space. *Research in Learning Technology, 24,* xx-xxx.

Miller, H. (2013). Investing in bad science. *Policy Review, 117*, 33-42.

Moreno, R., & Mayer, R.E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309-326.

Morrison, N.M., Burhnam, D., & Morrison, B.W. (2015). Cognitive load in cross-modal dual-task processing. *Applied Cognitive Psychology, 29*, 436-444.

Mythen, L., & Gidman, J. (2011). Mandatory training: Evaluating its effectiveness. *British Journal of Healthcare Management, 17*(11), 522-526.

Office of Personnel Management. (2013). *Employment and Trends: September 2013*. Washington, DC.

Ogden, R.S., & Jones, L.A. (2011). Modality effects in memory for basic stimulus attributes: A temporal and nontemporal comparison. *The Quarterly Journal of Experimental Psychology, 64*(7), 1354-1371.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105-119.

Pastotter, B., Weber, J., & Bauml, K.H.T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology, 27*(2), 280-285.

Peterson, K., & McCleery, E. (2011). Evidence brief: The effectiveness of mandatory computer-based training on government ethics, workplace harassment, or privacy and information security-related topics. *In: VA Evidence-based Synthesis Program Evidence Briefs; VA Evidence-based Synthesis Program Reports*.

Phelps, R.P. (2012). The effect of testing on student achievement, 1910-2010.

International Journal of Testing, 12, 21-43.

Preusser, M.K., Bartels, L.K., & Nordstrom, C.R. (2011). Sexual harassment training: Person versus machine. *Public Personnel Management, 40*(1), 47-62.

Putnam, A.L., & Roediger, H.L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition, 41*, 36-48.

Reynolds, M.E., & Givens, J. (2001). Presentation rate in comprehension of natural and synthesized speech. *Perceptual and Motor Skills, 92*, 958-968.

Rice, S., Geels, K., Hackett, H.R., McCarley, J.S., Schwark, J., & Hunt, G. (2012). The harder the task, the more inconsistent the performance: A PPT analysis on task difficulty. *The Journal of General Psychology, 139*(1), 1-18.

Riech, K.B. (2015). Psycho lawyer, qu'est-ce que c'est: The high incidence of psychopaths in the legal profession and why they thrive. *Law & Psychology Review, 39*, 287-299.

Roediger, H.L., & Karpicke, J.D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210.

Roediger, H.L., & Karpicke, J.D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255.

Rogalski, Y., Altmann, L.J.P., & Rosenbek, J.C. (2014). Retrieval practice and testing improve memory in older adults. *Aphasiology, 28*(4), 381-400.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 233-239.

Rosenthal, R. & Rosnow, R.L. (1975). *The Volunteer Subject*. New York: Wiley.

Rowland, C.A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432-1463.

Rowland, C.A., & DeLosh, E.L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory, 23*(3), 403-419.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Semb, G., Glick, D.M., & Spencer, R.E. (1979). Student withdrawals and delayed work patterns in self-paced psychology courses. *Teaching of Psychology, 6*(1), 23-25.

Shuyan, L., Kuschpel, M.S., Schad, D.J., Heinz, A., & Rapp, M.A. (2015). Differential effects of music and video gaming during breaks on auditory and visual learning. *Cyberpsychology, Behavior, and Social Networking, 18*(11), 647-653.

Simon, H.A., & Barenfeld, M. (1969). Information-processing analysis of perceptual processes in problem solving. *Psychological Review, 76*(5), 473-483.

Simonds, T.A., & Brock, B.L. (2014). Relationship between age, experience, and student preference for types of learning activities in online courses. *Journal of Educators Online, 11*(4).

Son, J.Y., & Rivas, M.J. (2016). Designing clicker questions to stimulate transfer. *Scholarship of Teaching and Learning in Psychology, 2*(3), 193-207.

The federal role in education. (2016, July 21). Retrieved from: https://www.2.ed.gov/about/overview/fed/role.html

Thomas, R.C., & McDaniel, M.A. (2013). Testing and feedback effects on front-end

control over later retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 437-450.

Thorndike, E.L., & Woodworth, R.S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*, 247-261.

Tompkins, L., & Ulus, E. (2016). 'Oh, was *that* "experiential learning"?!' Spaces, synergies and surprises with Kolb's learning cycle. *Management Learning, 47*(2), 158-178.

Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized assessment of reading performance: The new international reading speed texts IReST. *Investigative Opthalmology & Visual Science, 53*(9), 5452-5461.

Trumbo, M.C., Leiting, K.A., McDaniel, M.A., & Hodge, G.K. (2016). Effects of reinforcement on test-enhanced learning in a large, diverse introductory college psychology course. *Journal of Experimental Psychology: Applied, 22*(2), 148-160.

Tse, C.S., Balota, D.A., & Roediger, H.L. (2010). The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. *Psychology and Aging, 25*(4), 833-845.

United States Government Accountability Office (2012). *Federal Training Investments: Office of Personnel Management and Agencies Can Do More to Ensure Cost-Effective Decisions*, (GAO 12-878). Washington, D.C.: U.S. Government Printing Office.

Verhaeghen, P., & Salthouse, T.A. (1997). Meta-analyses of age-cognition relations in

adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122,* 231-249.

Weld, C. (2014). Listen to this! Utilizing audio recordings to improve instructor feedback on writing in mathematics. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies, 24*(6), 513-528.

Weng, P. (2015). Developmental math, flipped and self-paced. *PRIMUS: Problems, Resources and Issues in Mathematics Undergraduate Studies, 25*(9-10), 768-781.

Wetzel, C.D., Radtke, P.H., & Stern, H.W. (1994). *Instructional effectiveness of video media* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wheeler, M.A. (2000). A comparison of forgetting rates in older and younger adults. *Aging, Neuropsychology, and Cognition 7*(3), 179-183.

Wojcikowski, K., & Kirk, L. (2013). Immediate detailed feedback to test-enhanced learning: An effective online educational tool. *Medical Teacher, 35*, 915-919.

Wright, G., Shumway, S., Terry, R., & Bartholomew, S. (2012). Analysis of five instructional methods for teaching sketchpad to junior high students. *Journal of Technology Education, 24*(1), 54-72.

Zaromb, F.M., & Roediger, H.L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38*(8), 995-1008.

Appendix A: Full Assessment

The purpose of this test is to assess your knowledge regarding the USERRA act. The findings may be used to inform the revision of instructional delivery methods used in mandatory federal government training.

PART I: In-Presentation Test

True/False: For items 1-4, please select whether the statement is either "true" or "false."

1. True or false: USERRA guarantees the right to reemployment for reservists and National Guard members.
    a. True
    b. False

2. True or false: Accrual rates of 15 days per fiscal year apply to *active* duty training only.
    a. True
    b. False

3. True or false: Employees are required to use annual vacation or leave while on duty.
    a. True
    b. False

4. True or false: Under USERRA, in the event of an unsuccessful complaint resolution, an employee may receive no-cost legal representation in a district court.
    a. True
    b. False

Multiple choice: For items 5-8, please select the single best answer from the choices listed below.

5. Following alleged violations, complaints may be filed to _____.
    a. the Office of Personnel Management (OPM)
    b. the Merit Systems Protection Board (MSPB)
    c. the Dept. of Labor's Veterans' Employment Training Services

d. either the MSPB or the Dept. of Labor's Veterans' Employment Training Services

6. After the receipt of notification of upcoming service, employers must do all of the following, *except* _____.
    a. act on and approve the request
    b. treat the service as a paid leave
    c. administer USERRA waiver provisions
    d. inform the employee about entitlements and benefits

7. The "escalator principle" refers to _____.
    a. annual leave accrual
    b. the extension of healthcare benefits
    c. personnel exemptions due to short notice deployment
    d. the employee's advancement as if continuously employed

8. Following a period of military service longer than 181 days, employees have _____ days to apply for restoration of benefits.
    a. 15
    b. 30
    c. 90
    d. 120

PART II: Post-test (Transfer)

Multiple choice: For items 1-8, please select the single best answer from the choices listed below.

9. (1) The USERRA act *specifically* protects against discriminatory action on the basis of _____.
    a. citizenship
    b. sexual orientation
    c. military obligations

10. (2) Which of the following is *NOT* a valid reason for denying promotion following return from deployment according to the USERRA guidelines?
    a. Active duty service exceeds 90 days
    b. An employee was in a "developmental" position
    c. The employee cannot perform duties of the new position after reasonable training
    d. The employer decides to provide an alternate job if the returning service member is unable to carry out the duties of the new position

11. (3) The annual leave accrual rate includes the provision of up to _____ additional days for emergency service.

a. 12
b. 22
c. 32

12. (4) Which of the following provisions is *NOT* guaranteed under the USERRA statute?
    a. Prompt restoration to employment after service
    b. Assuring that an employee's office space does not change
    c. Ensuring that employees retain health and other benefits during service

13. (5) The provision of "reservist differential" pertains to a service member's _____.
    a. income
    b. health insurance
    c. leave accrual rate

14. (6) Mitul is a member of the Missouri National Guard, and is preparing to return to his job in two weeks. As Mitul's recent deployment was in response to *state* issues (specifically, flooding), his reemployment status is covered under which Title contained in USERRA legislation?
    a. 9
    b. 17
    c. 25
    d. 32

15. (7) In order to retain the right to restoration of employment following service, an employee must receive a discharge deemed "_____."
    a. punitive
    b. honorable
    c. either honorable or punitive

16. (8) Ruby is a Naval Reserve member, and has been deployed in Africa for the last three months. During that time, her previous civilian employer (the Department of State) has undergone a substantial restructuring at the local level, and her prior position of budget analyst was eliminated. Ruby is told that she has been placed with the Department of Commerce in a similar position at a nearby location. Has Ruby's employer acted in accordance with the USERRA act?
    a. No
    b. Yes
    c. Not sure

True/False: For items 9-15, please select whether the statement is either "true" or "false."

17. (9) True or false: Retirement falls under the protection of service credit USERRA provisions.

a. True
b. False

18. (10) True or false: In addition to current service member, USERRA also covers those who have *applied* for uniformed service.
    a. True
    b. False

19. (11) True or false: Title 10 status refers to National Guard members performing *State* duties.
    a. True
    b. False
20. (12) True or false: Alexis is a member of the North Carolina National Guard. She must miss work next week in order to participate in an annual fitness/skills assessment. Her absence is excused under the terms of the USERRA act.
    a. True
    b. False

21. (13) True or false: An employee must *always* provide advanced notice of deployment to an employer.
    a. True
    b. False

22. (14) True or false: When federal supervisory attorneys are deployed, their status as supervisory is temporarily forfeited.
    a. True
    b. False
23. (15) True or false: Expert witness fees are subject to inclusion as part of court awarded damages when employers are found to have willfully violated USERRA guidelines.
    a. True
    b. False

Multiple choice (scenarios): For items 16-20, please read the scenarios and choose the answer that best reflects your understanding of the material.

24. (16) Linda is a member of the National Guard and is currently a probationary employee with the Department of Justice. She is called away to provide federal assistance due to the aftereffects of a severe flood. Linda returns from deployment after a period of 18 days. Under federal guidelines, is Linda's employment status protected according to the USERRA act?
    a. No
    b. Yes
    c. Not sure

25. (17) Stuart is a full-time Department of Defense employee and serves in the Army Reserve. He is called away for service on short notice and is away from his DoD job for 45 days. Stuart was scheduled for promotion to an elevated position with a higher pay grade before being deployed. Upon his return, Stuart's supervisors inform him that he is still categorized as an employee in the previous position, stating that (due to his time away) he has not yet "earned" the promotion. In accordance with USERRA provisions, has Stuart's employer acted appropriately?
    a. No
    b. Yes
    c. Not sure

26. (18) Rosita is an attorney with the DoJ and currently deployed as a member of the National Guard in state service to provide hurricane relief to coastal communities on the gulf coast of Florida. Rosita learns from a superior that her deployment will end in five days. She calls her HR representative at the DoJ and verbally conveys her intent to return to employment following the end of her deployment. Has Rosita provided sufficient notification to her employer under the USERRA guidelines?
    a. No
    b. Yes
    c. Not sure

27. (19) While working as a full-time employee of the DoD, Malik, a Naval Reservist, was called into overseas service for a period of three months. During his term of service, Malik received a conduct-based punitive discharge. After returning stateside, Malik was informed that he would not be reinstated as a DoD employee. Has Malik's employer violated the USERRA act by not guaranteeing his reinstatement?
    a. No
    b. Yes
    c. Not sure

28. (20) Jordan files a complaint with the Merit Systems Protection Board (MSPB) alleging discrimination against her employer. She was wounded while serving in Afghanistan, and convalesced for a period of 18 months. Upon attempting to return to her old job at the DoJ, Jordan's supervisor informed her that the position had been permanently filled 6 months prior. Jordan's claim is not upheld, and she feels the situation was not handled properly. If she decides to appeal her case in District Court, should she reasonably expect her attorney(s)' fees to be covered under USERRA provisions?
    a. No
    b. Yes
    c. Not sure

Short Answer: For item 21, please describe briefly (in 2-4 sentences) what you would do in the following situation.

29. (21) You are a DOJ supervisor. Preston, one of your employees, has notified you in writing of an upcoming deployment. He asks you whether his period of service will be considered a paid leave, and if he is allowed to use accrued leave to cover that timeframe. How will you answer Preston's questions? Provide a brief justification for your response.

*You should let Preston know that his leave of service will **not** qualify as a paid leave under the USERRA act. Also, inform Preston that he is **allowed**, but not **required** to use is annual accrued leave during his period of service.*

Attitude Assessment

Multiple choice: For items 22-24, please select the single best answer from the choices listed below.

30. (22) An appropriate amount of content was contained in the USERRA training module.*
    a. Strongly disagree
    b. Disagree
    c. Agree
    d. Strongly agree

31. (23) The *format* of the training helped to facilitate my understanding of the material.*
    a. Strongly disagree
    b. Disagree
    c. Agree
    d. Strongly agree

32. (24) I would recommend this training module to others interested in learning about the USERRA act.*
    a. Strongly disagree
    b. Disagree
    c. Agree
    d. Strongly agree

Short Answer: For item 25, please provide additional information as to your overall impression of the USERRA training experience.

33. (25) Please use the space below to provide any comments regarding the USERRA training module (e.g., content, instructional method).