Fall 2018

Memorization And Mastery: Teaching Disciplinary Vocabulary In The Secondary Classroom

Matthew Granoff
Hamline University

MEMORIZATION AND MASTERY:

TEACHING DISCIPLINARY VOCABULARY

IN THE SECONDARY CLASSROOM

by

Matthew B. Granoff

A capstone submitted in partial fulfillment of the
requirements for the degree of Master of Arts in Teaching.

Hamline University

Saint Paul, Minnesota

December 2018

Primary Advisor: Jeff Fink
Secondary Advisor: Margaret McCarney, PhD
Peer Reviewer: Missy Johnson

To my wife and my parents,

who have learned the true meaning of "long-suffering."

Table of Contents

CHAPTER ONE

Introduction

I distinctly remember one of my professors advising my class, while I was first studying for my teaching license at Hamline, that we shouldn't take our first job at a charter school. I remember this so distinctly largely because I flagrantly ignored it and wound up with an amazing position at a truly exciting school, surrounded by exceptional colleagues – the sort of teaching job I wouldn't have expected to be able to land until much later in my career. I taught there for four years before moving away to follow my wife's career, and while I still love the school (even in exile), I do understand much more where my professor was coming from: as satisfying as I found my work, I did have to radically shift my understanding of pedagogy, both theory and practice, away from what I studied during my training.

On the plus side, my curriculum itself, though unusual, was fantastic: much better suited to my strengths than the sort of pure literature-based classes that I learned to teach in Hamline's MAT program. The school follows the Classical model of education, guiding students from Kindergarten to 12th grade through the three stages of Grammar, Logic and Rhetoric. Although I trained and licensed as an English teacher – excuse me, Communication Arts and Literature – I fell naturally and happily into primarily teaching Rhetoric. My classes were a joyful congeries of composition, public speaking, argumentation, philosophy and analysis, and I loved that I got to spend every day talking about the intricacies of language and communication and how to discover, articulate and effectively express a thought.

On the other hand, the Classical model (at least as interpreted there) also dictated that teachers use pedagogical strategies so far outside the window of mainstream public school approaches that not only did Hamline not cover them, but often gently discouraged their use. The emphasis in my licensure training was, to borrow a professor's phrasing, becoming a "guide on the side" rather than a "sage on the stage" – focusing, in other words, on the teacher as one who *enables* learning, for instance by creating an environment through careful scaffolding in which students can learn collaboratively and constructively. In contrast, my school explicitly wanted its teachers to present ourselves as experts and authorities in our fields. Some of the techniques it asked of me were a natural fit for my personal skill set – I've always been a storyteller/performer, for instance, so I found the School of Rhetoric's (grades 9-12) use of lecture-based classes more fun than difficult– but others rubbed the wrong away against both my inclinations as an instructor and, honestly, my preferences as a student.

One of those I found most difficult to successfully integrate into my classroom was the focus on extensive memorization. Many schools see "memorization" and "conceptual mastery" as distinct and even competing goals which must be balanced against each other; that school sees the former as effectively a prerequisite for the latter, and so students are given dozens or even hundreds of terms, dates and other technical vocabulary in the majority of their subject-area classes. My wife is a doctor, and as a medical student she somehow memorized literal thousands of terms; I struggle to master even a few at a time.

A brief defense and explanation of this emphasis is perhaps warranted at this point, given that different individuals find memorization everything from an unconscious

skill to a nearly unattainable one. One of the fundamental underpinnings of the Classical model is that, before you can have deep thoughts about an idea, you must have total familiarity with its basic building blocks. Elementary school students should memorize their times tables so that they don't have to *calculate* every time basic multiplication comes up in more advanced math; middle school students should memorize dates so they don't have to reconstruct historical chronology from scratch every time they learn something new; and good rhetoricians should memorize the figures of speech so, when confronted with a text, they can jump immediately to thinking about *why* the author made the choices they did, rather than wasting time puzzling out what those choices *are*. Moreover, this basic familiarity allows them access to deeper thoughts in other classes as well. You can't engage fully with Dante or Milton (both authors covered in the 10th grade literature course) without recognizing periphrasis or chiasmus on sight. On a more menial level, it's much easier to ace the AP English Language and Composition Exam if you're ready to look for anaphora and antithesis in the Rhetorical Analysis prompt. Though the task seems obscure and arbitrary to the students when they first encounter it, the grunt work of memorization in my class pays off in rich dividends for the remainder of their academic careers.

While my old school places memorization front-and-center of basically every subject area, even schools using much more commonly-employed pedagogies can benefit in this way from small integrations of the practice. For example, although my new school in California generally places far more emphasis on student-driven learning than my previous one, freshmen are still asked to memorize about two dozen literary terms. These are assigned in early September, and while we use the terms in classroom discussion all

fall, they are not formally tested until the semester final in December. The other teachers on my team rave about this approach, explaining that the final forces the students to actually *learn* the terms for real, and thus their thinking (and subsequently their discussions) in the spring really crystallizes. From an newcomer's perspective, it seems clear that rote memorization is equally important to our learning goals as it is to my previous school's, and it's primarily the (real but unspoken) cultural discomfort with explicitly *calling for* rote memorization that leaves us waiting for January to really get rolling with those goals.

All that said, I certainly have my own initial personal discomforts with this type of learning, which were borne out in the results of my first year trying to teach it. The sophomore Rhetoric course focuses heavily on style, learning both to recognize and to use it. Students do a lot of both analysis and their own writing in a variety of forms, and are meant to acquire a "toolbox" of terminology with which to articulate their thoughts, including a comprehensive familiarity with over 75 rhetorical devices and other terms. Or at least, that is the theory; the final exam demonstrated that many of them had significantly failed to master the admittedly arcane taxonomy (we went as deep as "epanalepsis," "antimetabole" and "paronomasia," none of which they are realistically likely to encounter again outside a Rhetoric classroom).

In a serendipitous turn of events, as I was in the process of rethinking my curriculum, our school indulged in the universal habit of administrators everywhere to, every few years, pick some Initiative to Roll Out. In the fall of 2014, we were introduced to the book that would be guiding our practice for the foreseeable future/until the next Initiative came around: *Make it Stick*, by Peter C. Brown. To my immense surprise and

delight, it turned out that this was actually a *really good book*. Although Brown has a habit of inventing new buzzwords where none were really necessary (e.g., "revisiting old material" is now "spiraling back"), the actual analysis and advice is based on extensive, well-cited research that, although it does contain very recent studies, is also grounded in work dating back multiple decades. I will be talking much more about this book, related work and other studies in the field in Chapter 3. For the time, however, we were "encouraged" to think of ways we could implement the principles of *Make it Stick* within the Classical model.

Further fortuitously, at this time I was also making my first attempt at writing my Capstone for my MAT degree at Hamline, and so looking for some sort of large-scale project I could tackle in my classroom and then report on. Considering all the factors together, and in consultation with my original advisor, I decided to focus on investigating how I could strengthen my students' acquisition of disciplinary vocabulary, i.e. the rhetorical jargon. I started doing research and, mid-year, assembling a plan of attack: an entirely new approach to a vocabulary-acquisition curriculum.

Unfortunately, I failed to complete my capstone on that attempt for entirely unrelated reasons, but I kept the project going regardless, refining its implementation as I gathered more research and adapted to the idiosyncrasies of my classroom and students, ultimately formulating the question, "how can I best teach disciplinary vocabulary in the secondary classroom?" To my initial frustration, I discovered that almost no research has been done on the specific subject of how *disciplinary* vocabulary is acquired by high school students—so little, in fact, that I'm still not sure whether "disciplinary vocabulary" is the "right" term to use when discussing technical terminology/jargon/subject-area

language in the first place. Ultimately, however, this scarcity of prior research opened up for me a really interesting opportunity to create a research-based curriculum from whole cloth. The curriculum presented in Chapter 4 of this capstone, rather than drawing from research and theory specific to its subject matter, is synthesized instead from research on multiple tangentially related topics. In other words, I see my work here as, basically, orienteering: I'm attempting to locate a new point on the academic map via triangulation from a few other known landmarks. It's a task that's both directly relevant to my classroom and surprisingly exciting intellectually.

CHAPTER TWO

Review of the Literature

Chapter Overview

The question of **how best disciplinary vocabulary can be taught in the secondary classroom** is one with many moving parts. First, of course, we must defend the premise that the direct instruction of vocabulary is a worthy goal to begin with; much of modern pedagogical practice can be seen as a reaction against the memorization-heavy practices in primary and secondary schools of the past. A look into current educational theory will help establish this groundwork, however, at which point we can move into to two key areas: how vocabulary is learned, and how it can be taught. Research reveals that this latter question can be subdivided further into two main components: independent student study and, somewhat surprisingly, regular and carefully crafted in-class assessments, to take full advantage of the somewhat counterintuitive "testing effect."

Education Theory

A lot of what goes on inside the classrooms of a modern charter school using the Classical model of education is quite distinct from those of the average public school. Even within Classical schools, there can be a wide variety of approaches; while my previous one focuses heavily on using Classical educational techniques such as dialectic (a sort of Socratic dialogue between the teacher and a single student) and memorization, others may read texts from the Western Canon while using much more mainstream pedagogical strategies. The theory behind creating my school, however, is that the older approaches are not only relevant and meaningful, but as effective at creating citizens of

the modern world as they were at creating citizens of ancient Athens. Very little has been written in the modern literature directly on Classical techniques, but there are a number of articles which discuss the long-term benefit of a number of individual strategies such as a focus on memorization, oratory and complex analytical tasks.

At the same ages where most American schools divide into elementary, middle and high schools, the school I taught at, a K-12 charter, splits into the Schools of Grammar, Logic and Rhetoric respectively. The interpretation of Classical education being used here is that, at a young age, students learn a whole lot of pure information before subsequently learning how to appropriately combine that information and draw conclusions from it, graduating finally to reasoning their way to independent conclusions. The first stage, Grammar, involves a degree of rote memorization that often shocks outside observers: by fifth grade, the students have memorized not just the US states and their capitals, but the presidents, the countries of the world, the water cycle, the cases and declensions of Latin, and the dates and major players of the War of the Roses. What often wins over skeptical observers is that the children *love* the work: teachers use a variety of mnemonics like songs and games, and the students soak up the knowledge. This is not, of course, an attitude which continues indefinitely: by the time they reach the School of Rhetoric (having passed through the analytically-focused School of Logic, where they learn how to piece together the various data points they've accumulated), most students are fairly tired of memorization and dismayed that, although they do now get to engage on a much higher intellectual level (for example, generating their own interpretations of and ideas about texts), they still aren't free from the expectations to learn vocabulary, dates and terminology.

It is at the secondary level, then, that the use of memorization as a pedagogical technique most requires defense. Certainly, it has been part of Classical instruction in rhetoric since the Renaissance. Joseph S. Freedman (1986), investigating copies of and commentaries on the works of Cicero used to teach rhetoric in the 16th and 17th centuries, identifies that it was divided into "theoretical and practical components" (p. 228). The "theoretical" are what we would now consider classroom work – lectures, direct or small group instruction – while the practical involved "memorization, written exercises and oral exercises" (p. 228). This hearkens back to rhetoric instruction in ancient Athens, which while technically a literate culture still valorized memorization and oral tradition – consider Socrates' dismissal of written speeches in the *Phaedrus* (Plato, 360 BCE) – and expected its orators to deliver their speeches from memory. In fact, in Book I of his *De Oratore* Cicero (55 BCE) quite literally canonized *memoria* as one of the five essential skills of rhetoric, along with discovery, arrangement, style and delivery.

Of course, demonstrating that memorization *was* an important part of rhetoric instruction is hardly sufficient proof that it ought to *continue* to be; it is hardly difficult to think of other practices from the Renaissance which have since been justly abandoned. "Rote memorization" in particular has acquired quite the negative reputation; in a paper in the *Journal of College Science Teaching* reflecting on her recent college-level biology course, instructor Cori Fata-Hartley (2011) argues strongly against lecture-format classes and rote memorization-based examinations, observing that far more students missed the "simple recall" question on her unit exam than missed the adjacent essay question on the same subject (p. 37). Although acknowledging that this was not a formal experimental context, Fata-Hartley attributes the disparity in performance to the fact that the

theoretically "harder" but more successfully-answered question "required more complex cognitive behaviors" (p. 37). This dichotomy is supported as well by a paper in the *Journal of Developmental Education* (Elder & Paul, 2010) which advocates deliberate instruction in critical thinking: "Without critical thinking guiding the process of learning," they argue, "rote memorization is likely to become the primary recourse" (p. 39). To these and other authors, "rote memorization" is used to mean, in essence, learning a sequence of words without *understanding what they mean*, thus preventing the student from *using* them to build further knowledge.

However, this understanding of the role of memorization in learning is deeply limited: even if we allow that rote memorization *is* learning words without meaning, it can still be an essential step on the road to deeper mastery. For an intuitive example, consider the multiplication tables: the goal of memorization is specifically so that, when asked "what is 7 times 6," the student does *not* think about the "meaning" of the question or do the calculation in their head, but rather spits out "42" without pausing for breath. Imagine how much more time-consuming basic algebra would be – not to mention personal finance – if those 144 most common operations required active thought every time one encountered them!

Research and theory bear out this intuition. One major approach is summarized well by Lowery (2001) in the *Journal of Chemical Education*, discussing philosopher David Ausubel's "assimilation theory" and how "human constructivism," the theory of education propounded by Joseph Novak which still very much informs best pedagogical practices, derives from it. The core idea behind both the abstract and the practical theories is that new understanding is built actively by the learner, and only *in connection* to pre-

existing knowledge and understanding. In considering memorization, Lowery does note that "meaningful learning stands in direct contrast to rote learning," but identifies the former as the "substantive connection" of new concepts to prior knowledge (p. 1108). In other words, it's not that memorization *itself* is inherently "unmeaningful," but that it can become so if the material to be memorized is not meaningfully grounded by the teacher. Considering the theory of constructivism further, it becomes clear that, while substantively-connected memorization might make the mental network broader rather than deeper, that broadening itself allows for much further construction to take place. This interpretation is borne out by van Gelder (2005), who explicitly connects the ability to think critically in a discipline to "mastering a body of knowledge… This means, in part, acquiring the specialist vocabulary" (p. 44). His argument in essence is that immediate familiarity with the proper names for different concepts allows you to use the referenced concepts with much greater fluency and nuance, and that "this improved insight is the basis for self-monitoring and correction" (p. 44), two fundamental aspects of critical thinking. Further, in their practical advice to science educators, Fisher, Grant and Frey (2009) observe that "solid science literacy instruction requires attention to vocabulary" (p. 184), before offering five strategies to ground memorization in prior knowledge and broader significance.

(Before proceeding any further down the path of theory, it must be noted that some have argued it is less a path and more a rabbit hole. In an absolutely fascinating literature survey for the *Annual Review of Psychology*, Roediger (2008) looks at studies of memory dating back to the 1800s and concludes that of the various theories that have been advanced to explain the "laws" or mechanics of memory, *none* of them withstand

rigorous empirical testing. For every explanation which holds true under some conditions and in some experiments, other studies can be found which directly contradict it. In other words, while our understanding of how memory (and thus learning) works has likely *improved* over time, we should at no point delude ourselves into believing it a solved problem. One particularly interesting wrench-in-the-works: a 1929 study by Symonds and Chase, republished in the *Journal of Educational Psychology* in 1992, found that students who were instructed to simply practice for a test several times before taking it, with no explanation of its relevance or importance, nevertheless scored better than students who practiced slightly fewer times but were given either extrinsic (reward-based) or (attempted) intrinsic motivation (e.g. priming them on the value of self-improvement or discussing the personal satisfaction to be found in mastering a skill). In other words, even our deeply-held modern belief in student engagement being the best route to achievement does not necessarily stand up against all empirical tests.)

This theoretical grounding for the importance of memorization does seem to be borne out by practical observation and study. Grove and Lowery (2012) used the theories of Ausubel and Novak to break 12 students in a college-level organic chemistry course into four groups based on their degree and type of engagement with the material. Although they saw distinctions between intrinsically-motivated "meaningful learners" aimed at a "sound, conceptual understanding" of the material (p. 204) and, on the opposite end of the spectrum, "indifferent learners" relying on rote memorization (p. 205), they observed that students in the first group *also* used "rote" techniques like flashcards; indeed, the central issue with the other students was that they *only* used rote techniques, "despite the knowledge that more meaningful ways to learn exist" (p. 205).

The issue was not with memorization itself, but how it was integrated into a broader system of learning. Kail (2008) reaches a similar realization in her reflection on teaching high school students to memorize Latin and Greek word roots: although initially reluctant to require any rote memorization, believing it "was not the type of activity that would encourage student engagement, interest, and exploration" (p. 63), she discovered that students were actually *excited* about the chance to "play with words" and "synthesize their knowledge" (p. 64). Ultimately, Kail observes that the memorization work in English led to benefits "across the curriculum" (p. 65) in subjects as disparate as math and biology.

To return to where this section began, modern scholars have observed distinct benefits from memorization when implemented in traditional, Classical contexts as well. Miller (2005) observes that, in Classical rhetoric, *memoria* involved not just memorizing terms but entire speeches, word-for-word, and designs a classroom activity in which students do the same for (teacher-curated) selections from great modern speeches. He notes that the act of rote memorization – seemingly outdated in an age with teleprompters – allowed students to have much deeper and more complex thoughts about not merely the subtleties of delivery but how to more effectively compose and structure their own oral arguments. Fritz and Weaver (1986) go further, arguing that instruction in all five of the canons of rhetoric also teaches fundamental critical thinking skills – *memoria* in particular, they argue, helps students grasp the subtle but critical differences between oral and written modes of communication.

To sum up, as long as "rote memorization" is understood to refer to a *method* of memorization, rather than a *purpose* (or lack thereof), theory and evidence strongly

supports its use as a pedagogical tool in the modern classroom. While memorization should perhaps never be an end in itself, when the material to be memorized is both grounded in prior context and fundamental to the construction of future understanding, it not only does not lead students to disengage with their learning but actively helps them to build new connections and develop advanced critical thinking skills.

Vocabulary Acquisition

Oddly, very little research – none that I was able to locate, certainly – has ever been conducted on the acquisition of disciplinary vocabulary: the terminology learned in science classes, or the names and dates of history, or (most relevant to this paper) the linguistic techniques of literature and rhetoric. So little has this area been considered, in fact, that I am uncertain whether any single standard term exists for it: I will continue to use "disciplinary vocabulary," but "technical vocabulary," "subject-area terminology," "academic vocabulary" or even "jargon" seem equally valid choices.

Fortunately, quite a lot has been written on other types of vocabulary acquisition, primarily in early childhood and in second language learning. No individual paper discussed in this section connects directly to my ultimate focus, but by looking at the aggregate of a variety of studies – including a few that focus on college-level learners – it is possible to draw some conclusions about general patterns in successful vocabulary teaching and learning strategies.

First, a helpful if not authoritative definition of (what I mean by) disciplinary vocabulary can be drawn from Chung and Nation (2003), who studied "technical vocabulary" in an anatomy and a linguistics text. They offer a four-step scale for categorizing words, moving from universal "function words" at Step 1 to words that have

a specific meaning within one discipline and are unlikely to be known or used outside it at Step 4 (p. 105). Step 4 clearly fits what most people would think of in terms of disciplinary vocabulary; however, they also categorize as "technical" their Step 3, of words which are in common usage but have an *alternate* meaning specific to a field, such as "chest" in anatomy (p. 105). Usefully, Steps 3 and 4 collectively encompass all the specialized vocabulary students are expected to master for the AP English Language (frequently referred to as AP Rhetoric) exam: unfamiliar rhetorical devices like anaphora and chiasmus, and familiar terms with specialized meanings like "natural" (as in sentence order) or "concrete" (as in diction).

That being established, we must now approach the far murkier question of how vocabulary acquisition actually *works*. While perhaps no longer the defining work of the field it was when first published, Craik and Lockhart's (1972) paper "Levels of Processing: A Framework for Memory Research" still casts a long shadow, and most work since positions itself in some kind of relation to their theory of *depth of processing*. The core of the idea is that words are learned "better" – generally measured in terms of duration of retention – when the learning activity involves engaging with the *meaning* of the word and mentally manipulating the underlying concept than when the word is presented free of context, as a "pure" memory task. This theory has been tested countless times, and frequently supported by empirical data. For instance, Brown and Perry (1991) had adult ESL students (native Arabic speakers) learn English vocabulary using either "keyword" techniques (inventing mnemonics based on the sound of the word), "semantic" study (using the word in context) or a combination of both (p. 662). They found that the keyword-semantic method produced results "significantly" better than the

keyword-only method, and "slightly" better than the semantic method – findings that they note are "consistent with the predictions made from the depths-of-processing theory" (p. 665). Similar results were produced by Ellis and He (1999) in their study of ESL students (Asian, primarily Korean) at Temple University. In this study, some students practiced words by carrying out instructions with the words in them (e.g. "put the cushion on the sofa"), others were allowed to ask questions or request clarification on those instructions, while a third group had to *create* such instructions given only a bank of labeled images. The researchers' intent was that the former two approaches – "input treatments" – would not engage the same depth of processing or conceptual work as the latter "output treatment," and their results bore out this belief: the output treatment group outperformed both input treatment groups in all three tests they administered (p. 296).

Some meta-analyses of the field also support the depth-of-processing claim, although others shed doubt on the magnitude of its significance. Laufer and Hulstijn (2001) fall into the former camp, going further and attempting to identify the characteristic attributes of learning exercises which require deep processing. Surveying the literature, they find three primary recurring factors – need, search and evaluation – and offer the term "involvement load" to represent "the combination of the presence or absence of [those] involvement factors" (p. 15). Checking this hypothesis against the published data, they find strong evidence that learning tasks across multiple studies with heavy involvement loads consistently led to greater retention; they also demonstrate that it can mesh on a theory level with the input/output theories tested (among other studies) by Ellis and He, discussed above.

In fact, few if any studies attempt to argue that depth of processing *isn't* a factor in vocabulary acquisition; rather, a number of them dispute the claim (made with particular transparency by Laufer and Hulstijn) that it – or its various derivative theories – is the *most important* factor. Consider Khoii and Sharififar (2013), who studied 38 native Iranian students enrolled in English Translation at Islamic Azad University in North Tehran. Over a four-month period, half of the subjects studied English vocabulary using rote memorization – which the authors specifically define as "repeating [material] over and over again until it is memorized" (p. 202) – while the other half used "semantic mapping," a visualization strategy that requires students to identify conceptual relationships between words for the purpose of "deepening understanding of a text and creating associative networks" for their vocabulary (p. 203). The students were tested at the end of the four months, and to the authors' surprise they found "no significant difference" between the two groups (p. 206). Both groups increased in vocabulary mastery from where they had started – both learning techniques were effective – but neither one was noticeably *more* effective than the other. Morin and Goebel (2001) found related results in their study of native English speakers taking their first semester of Spanish in college. All students took part in "communicative activities" (e.g. group conversations) to develop vocabulary, while half of the sections also used semantic mapping heavily (p. 12). Fascinatingly, while students in the semantic mapping group *felt* more confident and comfortable with the vocabulary, they consistently did not actually score any higher on the assessments. In an amusingly defensive presentation of these results, the authors note that, while both groups *knew* the same number of words,

"learners in the semantic mapping group did remember having at least heard more items than their peers in the vocabulary activities group" (p. 14).

Martin-Chang, Levy and O'Neil (2007), studying 2nd graders, report an unexpected variation on this theme. Some students learned new words within the context of narratives or other longer text passages, while others studied them in isolation (i.e. on flashcards). The context groups consistently scored between 2 and 10% higher in accuracy than the isolation groups on both the immediate assessments and after a 1-week period (to test retention), giving weight to the significance of deeper processing. However, comparing the initial acquisition scores to the retention testing scores, it turns out that neither group declined significantly over the period. In other words, although pure, context-free rote memorization was slightly less effective than context-grounded learning in terms of how *many* words could be taught over the teaching period, the two approaches were equally successful at creating longer-term retention of whatever words *were* learned.

To sum up: there are a lot of different ways to teach vocabulary, some of which are better than others much of the time except when it's the other way around. In other words, seeking the "best" method of instruction is a fool's errand. Let us therefore focus instead on attempting to catalog those methods which have shown any consistent success.

One theme which seems inarguable is that "direct" vocabulary instruction – the instructor introducing and/or teaching the words out of any immediate context – is distinctly powerful. Stahl and Fairbanks (1986), in a meta-analysis of studies on vocabulary acquisition, determined that direct pre-teaching of unfamiliar words before students read a passage containing them significantly increased student comprehension of

the passage. Adapting this idea to the specific case of technical scientific vocabulary, Brown and Ryoo (2008) observed that direct instruction that began in "vernacular" language before transitioning into jargon produced significantly higher gains in students compared to those who were instructed in technical language from the beginning: while both groups gained proficiency, the former students' scores improved by 12% *more* than the latter (p. 544).

Another recurring theme, and another interesting twist on the depth of processing theory, is that *metacognition* – teaching students to think actively about their own vocabulary-learning strategies – can be a major factor in student success. Mercer (2005) borrows Norbert Schmitt's division of strategies into "discovery" and "consolidation" (p. 25), and lays out a suggested plan for classroom use that begins with a 3-stage process of student self-reflection. Key to both the discovery and consolidation components are students interleaving direct study of vocabulary with observation and discussion with classmates and the instructor of successful and unsuccessful strategies. This aligns with the literature survey by Blachowitz, Fisher, Ogle and Watts-Taffe (2006) – of the three components of good vocabulary instruction they argue have emerged by consensus, the first and third both relate to active student awareness of the process of word learning. The second component, meanwhile, advocates for a combination of direct, "intentional" teaching of words with context-rich opportunities for practice in their use (p. 527). Moreover, a detailed case study of non-native English-speaking students in a graduate theology program (Lessard-Clouston 2008) finds all of these strategies in practical use. The students, learning such advanced and abstract terms as "theodicy" and "soteriology," actively considered the strategies which had been successful for them in previous courses

before employing a combination of group discussion, context-heavy reading, and even rote techniques like repetition and flashcard use.

One final tactic for vocabulary learning which does not fit neatly into any of the previous frameworks is what's called the spacing effect. In short: study sessions tend to be significantly more effective at encouraging long-term retention when they are spaced out over longer periods of time rather than closely clustered. A potentially definitive source on this topic is Cepeda, Pashler, Vul, Wixted and Rohrer's (2006) meta-meta-analysis of surveys of studies: noting that many authors, in testing the phenomenon, had not controlled for how *much* teaching was done at later interval points, they retroactively control for that factor and find overwhelming evidence in favor of the spacing effect. The implications for classroom teaching are obvious: vocabulary instruction cannot be a one-time thing if long-term retention is a goal.

Considering all the data, then, we find that there are a number of theories explaining the psychological mechanisms behind vocabulary acquisition which each appear to account for some portion of it, and a small variety of demonstrably effective strategies for effective teaching and studying. Perhaps the variation is due primarily to variance in student learning styles; in any case, the take-away is that while there is no single Best Practice, there *is* a healthy amount of evidence distinguishing better practices from worse ones.

Effective Flashcard Use Methods

Having reviewed what is known about the most successful strategies for teaching vocabulary, the next obvious field to investigate is the strategies students can use for independent study of same. In practice, this quickly became a survey of the effective use

of flashcards, for two reasons. First, a number of the studies on vocabulary acquisition used flashcard programs, so it seemed an important lead to follow. Second, flashcards are already one of the educational tools in widespread use in my school, and I wanted to either find confirmation that this was a productive strategy or, if not, identify an alternative. The research is overwhelming, however: flashcards, *used properly*, are an incredible student-centered tool for memorization in a wide variety of contexts. The following studies investigate a number of variables, allowing the careful reader to identify the most significant ones and plan a curriculum around those.

To begin with, it is clear that the same "spacing effect" discussed previously in the context of teacher instruction also applies to student study. This was tested directly by Kornell (2009) by having university students (in the context of a psychological experiment) learn a set of 20 words, over four study sessions, using two different methods. One group studied the entire set of cards every day, while the other used a set of 5 different cards each day (although spending the same total amount of time studying). The spacing effect would predict that the former group would have greater retention, since the space between seeing each word would be significantly extended, and this is in fact what the results found: performance in the spaced condition exceeded that in the "massed" condition by over 30% (p. 1304). This held true even in a follow-up experiment where both groups were given a final cumulative review session of all 20 words before the assessment. Spaced practice greatly impacts long-term retention.

Interestingly, however, the same experiment also found that student *perception* of their own learning was not aligned with their actual success – in other words, that the flashcard method they thought worked better was actually less effective. Another paper

by Kornell and Bjork (2009) further investigates this unconscious bias in judgment of learning, this time by asking students after one flashcard study session to predict how much their test scores would improve after 1, 2 or 3 subsequent sessions. Regardless of which group they were assigned to, students consistently predicted that their scores would rise minimally – no more than 15% – or even not at all. However, their actual mastery of the material soared by over 60% over the course of the experiment. The big takeaway for classroom teachers seems to be that spaced flashcard review *works*, even if students do not perceive it to be having any effect on their own learning.

Moreover, it works for all students regardless of initial ability level, as reported by Albers and Hoffman (2012). Working with an admittedly small sample size of 3 third grade ELL students, the researchers tried a number of different interventions including flashcard drills between assessments of the students' abilities to identify words that correctly completed sentences (words circled correctly per minute or WCCM). The three students had baseline WCCMs of around 4, 5 and 7, but over the span of 14 study sessions their scores all showed significant (if not continuous) growth, ultimately rising by between 6 and 12. Clearly, there is distinct value in a sustained program of flashcard use.

One further attribute of successful flashcard implementation relates to the composition of the "decks": an approach called Incremental Rehearsal (IR) has been shown to be significantly more effective than the "traditional" method. MacQuarrie, Tucker, Burns and Hartman (2002) outline the differences. The Traditional Approach starts with a deck full of unknown words, and drills until all are mastered (defined as getting them right three times), while IR starts with mostly known words, slowly rotating

in new terms and rotating out mastered ones so that the ratio of known to unknown in the deck is kept roughly steady throughout the practice. In both third and seventh grade students, IR both led to more words being learned in the initial study period and produced significantly greater retention over a 30-day period.

Of course, per the previously-discussed difficulty trying to pin down any hard and fast rules for how memory works, some researchers disagree with the findings in favor of Incremental Rehearsal. Volpe, Mulé, Briesch, Joseph and Burns (2011) object to studies like MacQuarrie et al. on the basis that they do not control for time spent studying with the various methods (a flaw the earlier authors do acknowledge; they defend their results by pointing out that the proportional gains from IR over TA are significantly greater than the proportional difference in study time). Volpe et al. ran a study on IR versus Traditional Drill (effectively the same as TA) where total study time between methods is kept even, and found that, although both produced increases in mastery, neither consistently outperformed the other. Of course, the potential issue with these results is that their sample size was 4, compared to the 51 students tested in the 2002 study.

Overall, the takeaway seems to be that flashcards, in general, clearly are effective at helping students study vocabulary, and while Incremental Rehearsal may be more effective for many students, for others it may not exceed the benefits of time and effort spent on more traditional approaches. In fact, there appear to be many aspects of flashcard use which are equally equivalent and can be safely left to student preference, such as format: Grillo and Decker (2013) observed that high school biology students showed equal success using paper flashcards as using digital programs like Study Stack. In the end, as we will see in the next section, what may matter much more than how one

sets up a flashcard study session is how one approaches and uses each individual flashcard.

Testing and Retrieval Practice

One of the most fascinating results to come out of the rigorous, scientific study of educational psychology is also one of the most counter-intuitive: *testing* a student on material learned is far more effective at creating long-term retention than having that student *study* the same material. This phenomenon is called the Testing Effect, and its lack of impact on modern pedagogy is surprising given its strength and consistency. Below are reviewed a number of studies of many different individual variables, from the aggregate of which we can consider what general plan for classroom (rather than laboratory) implementation is likely to best reap the benefits of the effect.

The core mechanism behind the testing effect is "retrieval." Although the psychology underlying *that* is still unknown, it has been verified repeatedly that the attempt to recall previously-learned information, to use a metaphorical description, strengthens the mental pathway involved and allows for it to be travelled more swiftly and confidently in the future. This is what makes flashcards so effective as study tools when used correctly: if the student looks at one side and *tries to remember* what's on the other side before flipping it to see, the act of trying solidifies the memory for future attempts. It really is the act of testing itself that is responsible for the effect; Haynie (1997) studied whether anticipation of a high-stakes test would have a comparable impact on long-term retention as actually taking one, and found that not only did the group which was tested significantly outperform the group which was merely told there would be one, but that the latter group showed barely any gains over a third group which was

neither tested nor told they would be. Similarly, Carrier and Pashler (1992) determined that replacing the test with equivalent, extra exposure to the material did not have an equivalent long-term effect on retention – retrieval, they concluded, "has beneficial effects for later retention above and beyond the effects due to merely studying the item" (p. 639).

The testing effect is more complex than these studies might suggest, however, and can even lead to such counterintuitive results that failing a test – even taking a test on advanced material before even being first introduced to the material – can still have long-term retention benefits. Much of the work on this phenomenon has been done by Nate Kornell; both Kornell, Hays & Bjork (2009) and Richland, Kornell & Kao (2009) report studies in which students took tests deliberately designed to be failed (on very difficult medical texts, obscure vocabulary words or, in some cases, purely made-up "facts"). Obviously, effectively all students failed these initial tests, but when they were retested after being given study time, students who had been pre-tested scored meaningfully higher than those who had simply been given more study time. In other words, paradoxically, retrieval practice can even solidify mental pathways that don't fully exist yet.

Of course, the testing event itself does not have to be the end of the testing effect: follow-up correction can also significantly impact long-term retention. Pashler, Cepeda, Wixted & Rohrer (2005), looking specifically at adult subjects learning foreign vocabulary words (selected for this study based on their unfamiliarity), found that when students who got a question incorrect were immediately shown the correct answer, two benefits followed. First, their scores on the subsequent exam rose significantly; second,

this new mastery, although it did decay over the following week, did not do so at a greater rate than that of any of the other groups. In short, immediate correction of wrong answers provided both immediate and long-term benefits.

Hays, Kornell & Bjork (2013) refine this idea further, discovering that feedback about missed items is most effective when presented immediately following the failure compared to when it is delayed. Their theory is that the testing event "primes" the mind, making it more receptive to correction (p.294); a reasonable analogy might be training physical technique, where if errors in form are not corrected quickly then further practice can simply solidify the poor habits. Unlike that analogy, however, students who receive immediate correction often do not *perceive* that they have actually learned anything. Kornell and Rhodes (2013), in further testing of the impact of immediate feedback, asked students to rate their confidence after each question that they would get it correct (a metric called Judgment of Learning or JOL). Although students receiving correction after errors showed massive improvement on subsequent tests relative to students who received no feedback, both groups had roughly equivalent JOLs. This loops back around, in a way, to the findings about student perception of the value of flashcard drilling, and strongly suggests that there are many ways in which what is best for ultimate outcomes may not be perceived as beneficial or even relevant by the students themselves – an especially tricky paradox for teachers trying to build their students' intrinsic motivation.

While the work of Kornell and others provides fascinating insights into the abstract psychology behind the testing effect, other researchers have investigated instead what sorts of tests and questions can provide the greatest benefits. For one, unsurprisingly, questions which access greater depths of processing are more effective

than more surface-level questions. Craik and Tulving (1975), in a follow-up to the original 1972 paper on the subject, tested various attributes (e.g. response speed and later recall) of responses to questions of varying complexity about words. At the shallower end were questions about upper vs. lower case; at the deeper were questions about appropriate vs. inappropriate use in a sentence, and the latter type of questions led to nearly 60% improved recall vs. the former on subsequent tests. Supporting this hypothesis, Butler and Roediger (2007) tested students in a simulated classroom setting by, following three lectures, having them either review a summary of the lecture or take either a multiple-choice or short-answer quiz. Students in the last group – who had to use much greater depth of processing – outperformed both other groups by nearly 10% on a final assessment. Further support comes from Roediger and Karpicke (2006a), who had their subjects read a prose passage before testing their recall of the general ideas from the passage (rather than a more concrete task like recalling word pairs). While this experiment did not directly compare different depth-of-processing levels, it did find that the testing effect still holds even with more general, abstract recall tests: retention dropped significantly less for the tested group than the one given more study time instead.

Going one step further, Wetzels, Kester, van Merriënboer & Broers (2011) determined that the testing effect can aid student learning even when almost entirely undirected by a teacher. Specifically, students with a prior knowledge base in a topic, when asked to simply take notes/review their knowledge on paper, showed significant gains in mastery relative to students who used purely oral or visual (reading) methods of review. Even without a specific prompt, the act of writing engages recall sufficiently to

impact long-term retention (although it must be noted that this method was not effective for students who did not already have sufficient prior knowledge).

Finally, one benefit of testing not directly related to the testing effect was articulated by Soderstrom and Bjork (2014). Taking the standard experiment for investigating the testing effect, a series of study and testing events, they interposed before the final assessment a period of self-paced and self-directed study, and discovered that students who had taken an interim test both focused on those areas in which they knew they needed the most review, and (more surprisingly) used effective study *strategies* at significantly higher rates than students who had only been previously given study periods. In other words, the role of testing in increasing student self-awareness and metacognition about their own learning appears to lead directly to more effective and more successful student-directed study in the future.

To wrap everything up, the research strongly supports several key ideas related to this study. First, there is clear value in a memorization-focused approach to mastery of technical vocabulary; second, that mastery is best reached through a combination of deep cognitive processing and more rote study and rehearsal. In terms of practical strategies for achieving that mastery, teachers ought to establish both a clear and directed program of independent flashcard study and a regular and carefully planned series of assessments interleaved with that study program. In the following chapters, we will examine what such a curriculum might look like in a modern (Classical) high school classroom.

CHAPTER THREE

Methods

Overview

Direct instruction on and memorization of disciplinary vocabulary is useful to students on multiple levels. Beyond allowing access to deeper levels of critical thought (Van Gelder, 2005), learning the technical terminology of many advanced high school subjects is a tremendous boost to success in those subjects' various Advanced Placement (AP) exams. Even more specifically, within the Classical model of education, memorization of terms and even of entire texts is a core element. The guiding question behind this capstone is how disciplinary vocabulary can best be taught in a secondary classroom; while the previous chapter was a broad survey of the literature relating to that question, in this chapter I will be extracting from that literature the key theories that guided my design process, along with establishing the context for the curriculum I have created.

Curricular Framework

In this project, I do not employ any single, central model for my curriculum; rather, I draw on several of the dominant theories which emerged in the previous chapter. The goal is to design lessons and other instruction to take fullest advantages of the scientific understandings of memory and learning.

That said, if there is a single text which provided an overarching framework for my design, it is *Make it Stick* (Brown, Roediger & McDaniel, 2014). Subtitled "The Science of Successful Learning," it actually cites several of the same studies as I did in the previous chapter, with an eye for adapting the results into general guidelines for classroom teachers. While I do not follow their five-part model of spaced retrieval practice, generation, interleaving, elaboration and calibration with perfect fidelity, their general approach to instruction of putting *retention* as the ultimate goal of classroom instruction has been influential both to me and to my school in general.

Specific theories from the research which I will have attempted to capitalize on in designing my curriculum include:

Depth of Processing. As first laid out by Craik and Lockhart (1972), the "depth of processing" theory claims that length of retention is directly correlated with the "level" of mental engagement with the material. While this is not directly connected to the more familiar Bloom's taxonomy, the latter provides a useful analogy: depth of processing argues, in essence, that the "higher" on Bloom's taxonomy the student is working, the more completely and longer they will remember the material. Although depth of processing has not been universally supported by experimentation – understandably, perhaps, given the difficulty of precisely determining internal mental processes – the vast majority of research does seem to agree that the fundamental correlation exists.

The Spacing Effect. As summarized in Cepeda, Pashler, Vul, Wixted and Rohrer's (2006) meta-analysis, the "spacing effect" describes the phenomenon in which information is retained better and longer if the learning sessions do not occur in immediate sequence but are spaced out over larger periods of time. While not every

attempt to replicate this effect was successful, depending on individual or environmental factors, when those confounders are controlled for then the evidence overwhelmingly supports the theory.

The Testing Effect. This is perhaps the most important theory guiding my design process. Beyond *Make it Stick*, the clearest summary of the theory comes from Roediger and Karpicke (2006), who find evidence in everything from psychology lab studies to "live" classroom environments to support the claim that the testing of knowledge directly leads to stronger retention of that knowledge. Kornell and other researchers have dug into what processes drive this phenomenon, and the consensus appears to be what *Make it Stick* summarizes as "retrieval practice": the act of *trying* to recall knowledge, whether or not it succeeds, strengthens the mental pathways which allow for smoother recall in the future.

Intended Audience

I have tackled two main tasks with this design project: first, laying out a generalized curriculum model for teaching disciplinary vocabulary which can be easily adapted to more or less any secondary classroom; and second, providing a specific adaptation of that generalized model to a Classical Rhetoric classroom. The first part is much more likely to be useful to a randomly-selected reader of this paper, but will by its nature need to be somewhat vague. By also including a specific potential implementation, I hope it can be used as a model for teachers in other disciplines.

For the generalized model, the intended setting is a high school classroom – I can see it being useful in an AP English Language class, various sciences or even history, but that is less an exhaustive list and more what comes immediately to mind. Any class

whose curriculum includes learning a lengthy list of disciplinary vocabulary could conceivably implement this plan. In particular, it is not limited by any class sizes a teacher is likely to encounter. As this plan is intended to supplement a primary subject-area curriculum (or perhaps replace those portions of it dedicated to vocabulary instruction), focusing purely on one specific target, it is designed to require as little additional grading time from the instructor as possible. Thus, time required for implementation should not scale up significantly with larger classrooms. Moreover, it does not require any time spent by students outside of the classroom – although extra work would almost certainly help, the curriculum is designed to work purely within the normal school day.

The one group of students I suspect this curriculum would specifically not work for is English Language Learners. Although I will be drawing from studies of ELL students in developing my design, and so individual activities may prove useful, since acquisition of academic language necessarily follows conversational language, some of the strategies may not be accessible. That said, I am not an expert on ELL instruction, and it is very possible that my understanding of those student needs may be inaccurate.

For the specific example I have developed, teaching the disciplinary vocabulary of rhetoric, the intended setting is a modern Classical high school rhetoric classroom. Based off my own previous school, what I have in mind is a classroom of 15-30 sophomores, majority white but with representation of all significant minority populations. The one difficult-to-measure factor that I believe impacts that student body is that, as my school is a charter, all its families have specifically chosen to enroll, which suggests active parental involvement in their children's education. While the specific

impact of such involvement on student outcomes is a whole other dissertation at the very least, it seems important to acknowledge that my mental model of students for whom I am writing this curriculum likely receive an above-average amount of it.

Design Process

I split the design process into four main chunks: initial instruction, follow-up lessons and activities, independent student work, and assessment. These have all been informed heavily by the theories encountered in my research. Both initial instruction and follow-up lessons are driven by the depth-of-processing theory, along with general guidance from *Make It Stick*. Independent student work revolved around designing a specific set of instructions for flashcard design, use and practice – the goal was to take the principles of spacing and retrieval practice and turn them into an easy-to-follow plan for students. Assessment design also centered on leveraging the testing effect and retrieval practice to the fullest extent possible, while also considering the utility to the teacher in terms of both minimizing the effort of implementation and providing formative feedback to guide future instruction.

Human Subject Review

As this is a curriculum design capstone project, I did not conduct any study of human subjects – while I "tested" variations of this design on my past classes, this was done at the behest and with the approval of my school, rather than for research purposes.

Summary of Chapter

To answer my research question of how best to teach disciplinary vocabulary to high school students, I have designed a curriculum based on several theories drawn from education, memory and psychology research and guided by the ideas from *Make It Stick:*

The Science of Successful Learning. I designed both a specific curriculum that might be implemented in a rhetoric classroom at my school, and a more generalized version that could be adapted for (theoretically) any secondary classroom. This design process considered the curriculum in four separate elements, divided by function, rather than presenting a day-by-day lesson plan. In the next chapter, I tackle those elements one by one, demonstrating how a practical method of instruction can be designed based on the underlying theories.

CHAPTER FOUR

Curriculum

Overview

In this chapter, I will design a complete study curriculum, designed to be used as a supplement to any independent unit featuring new disciplinary vocabulary, which best responds to the research and employs the techniques discussed in Chapter 2 of this paper.

Introducing the Disciplinary Vocabulary

Since, as discussed, a critical component of meaningful use of rote memorization is that said memorization not be the end goal of a curriculum, it is critical from the very first introduction of the disciplinary vocabulary to be learned that the terms and definitions be connected to and/or grounded in broader frames of thinking. In practice, this is what most modern teachers do anyway, but is perhaps worth reinforcing: students should receive more than an isolated list of terms and definitions. After all, the goal of rote memorization is speedy recall, to facilitate higher-level thoughts and discussions.

In my rhetoric classroom, what this looks like is that the initial introduction of every new figure of speech is accompanied by both a rigorous, formal definition and, side-by-side, a "common language" one. The former is the memorization goal; the latter serves both as partial "translation" and as link between the new concept and prior knowledge. Following this, I provide multiple examples of the figure of speech "in action" – since our long-term learning goals are to recognize the figures when we see them and to be able to interpret their rhetorical impact on an audience, I can use these examples as practice for both. With each example, there are two questions: "How is this

an example of [term *x*]?" and "How does it change the way we read and respond to the sentence that the author used [term *x*] as opposed to a more 'standard' phrasing?"

I don't think there's any single appropriate method of adapting this approach to other subject areas, and I would not presume to prescribe one. All I would emphasize – and I strongly suspect, again, that most teachers do this already – is that new terms should be introduced with linkages both to prior knowledge and to the formal academic context that will be the larger focus of the curriculum.

Reviewing and Reinforcing the Vocabulary

A cornerstone of this study curriculum is the careful and deliberate use of flashcards, and so it is important to properly prepare and set up the students to use them "correctly." One of the first steps I would recommend is to actually "sit down" with the students and spell out the theory and science behind all the "rules." I see this as reaping multiple benefits. First and perhaps most importantly, it should dramatically reduce the sense of arbitrary micromanagement. High schoolers, in my experience, are hyper-sensitive to anything that feels like "busy work," and flashcards (and, indeed, rote memorization in general) tend to set off bored eye-rolling unless very carefully introduced and positioned within a broader, purposeful plan. Thus, before even assigning the cards to be made the first time, I would speak about much of what has been discussed in previous chapters of this capstone: the Testing Effect, the framing of flashcards as retrieval practice, even the high-level perspective of thinking about disciplinary vocabulary like times tables – data points whose memorization enables and expedites higher-level thoughts.

In terms of constructing the cards themselves, the critical question for the teacher to ask is: which way is most important for the retrieval to go? In other words, is the goal for students to be able to produce the correct term when looking at or thinking about what the term describes, or to be able to produce the meaning of the term when encountering it? Traditional, non-disciplinary vocabulary instruction, for example, favors the latter: when learning new English words (assuming a native English-speaking student), the goal is to remember what they mean when one sees them in an unfamiliar text without having to grab a dictionary. With rhetorical devices – and my instinct is that this is generally true for most disciplinary vocabulary – the teacher has the inverse goal. I have trouble imagining a situation in which one of my students would be called upon to remember "what anthimeria means," but when they see a sentence where a word is used as the "wrong" part of speech, I want them to be able to say, "this author has used anthimeria." So, the flashcards should have the term on one side, and a definition that the student understands[1] on the other, and when student reviews the card it should always be by looking at the definition side and trying to recall the unseen term.

Teachers in other fields should dictate the form of flashcards based on their answer to the above question. If a history teacher wants students to remember the date an event happened, then they should look at a side with the event named on it and try to recall the year written on the other side; if a biology teacher wanted students to remember the defining features of a particular phylum, then they should look at the phylum name and try to recall the data points written on the obverse.

[1] In practice, since the quizzes will be based on the "formal" definitions, I expect many students will simply use those for the sake of expediency, but they should be reminded that rote memorization is not the same thing as *blind* memorization.

In addition to prescribing the method for using individual cards, teachers should also manage to the extent possible students' patterns and habits of study. Obviously, much of this is far beyond our ability to control, but given how firmly the research supports the idea of regular retrieval practice leading to long-term retention, teachers pursuing a jargon-heavy curriculum would be well advised to set aside some class time for flash card review at regular interviews. How much and how often may be somewhat arbitrary, but considering the length of the "study periods" from many of the studies reported on in Chapter 2, I would suggest 5 minute-long blocks of review time, at a minimum of at least a few times a week. Generally, the research suggests that, if more time is needed, it is more effective to increase the frequency rather than the length of these study blocks. My classes are 50 minutes long; mastering this terminology is important enough that spending 10% of that time on flash card review feels genuinely appropriate. Students should, of course, be strongly encouraged to supplement this time with further review outside of class hours, but at least this way some amount of regularity can be guaranteed.

During that time, students should use the methods of review discussed above and in previous chapters: look at the "front" of the card, attempt for a second or two to remember what information is on the "back," and then immediately flip the card over to check. Cards should then be shuffled back into the stack, whether or not the student remembered correctly. A card should not be removed merely because they "got it right," or even if they've remembered it successfully several times in a row; a card should be removed only when the student knows it so well that they don't even have to make a

conscious effort to recall it. Again, the analogy of the times tables is a useful reference point: the goal is not merely accuracy but speed.

I have one final thought on flashcard use, which I heard very recently from a colleague, makes a tremendous amount of intuitive sense and is borne out by both our practical experiences, but which I have not found specific research to support. There is concern that, when students work independently with flashcards, what they are "actually" memorizing is as much the specific *visual* of the card as the underlying concept represented thereon. This isn't to say that students are deliberately doing this to cheat or anything, merely that, when they see the exact same card face every time they practice retrieving a piece of information, the sight of the words becomes as associated with the information as their meaning. To counteract this, rather than working individually, students should work in partners, taking turns reading cards aloud to each other. Ideally, they would do more than just read the exact same words every time, but challenge themselves to find new ways of "cluing" each term every time one comes up, so that there is no pattern in the retrieval prompts *other than* the core understanding which is the associative goal. Again, in our classes at my school, we have seen much stronger results with this partnered, vocal approach than with solo, silent work, but it is entirely possible that this effect is merely a weird artifact of our particular situation and/or student body, so take this final paragraph with a grain of salt.

Assessing Mastery of Disciplinary Vocabulary

In designing a course of assessment, I've tried to balance efficacy with efficiency – having tried a few different approaches informally in my classroom, the plan described below meets the core educational goals while requiring as little of the teacher's time as

possible. This latter requirement may seem inappropriately selfish, but given both the research discussed previously that suggests that *frequent* tests are significantly useful in creating long-term retention, and the conventional wisdom that frequent, low-stakes formative assessments are crucial for guiding responsive curricula, the workload involved in writing, administering and grading each assessment is absolutely a necessary consideration. After all, the goal is for this study curriculum to *supplement* a pre-existing course, not replace it entirely.

The format of assessment I recommend is as simple as possible: a biweekly, ten-question quiz which provides the formal definitions of disciplinary vocabulary and asks students to write the correct term. Both the schedule and length can be easily altered; considering my course in specific, which meets every other day and dedicates 5 minutes of each meeting to flashcard review, 10 questions biweekly means that every term on my complete list of ~80 can be quizzed multiple times over the year.

Question Composition

I would use precisely the same question for each vocabulary word, every time it appears on a quiz. This directly connects the assessment to the practice (flashcards), and creates a predictability that allows students to focus on pure memory rather than interpretation. My assumption is that there are other assessments, elsewhere in the broader curriculum, which ask students to engage in higher-level thinking involving the terms; these quizzes are laser-focused on memorization alone.

One efficiency advantage of using the same questions is that it allows the teacher to create a "question bank," enabling *much* faster creation of new quizzes. Once you've decided which terms need to be tested, you can simply copy and paste the given questions

from your bank into the quiz template rather than having to compose or even type them from scratch. This allows me – having gotten used to this system after testing variations for some time – to do the actual *assembly* of a quiz (distinct from design, which will be discussed below) in something like 2 minutes, maximum.

Grading and Returning Quizzes

One obvious benefit to writing quizzes whose answers are all single terms is that, with the aid of an answer key, a single quiz can be marked and scored in literal seconds. I recommend keeping the marking as simple as possible: a slash through missed questions, and a score at the top (perhaps in a distinctively colored ink, if you're concerned about post-facto student alterations). In addition to making life easier for the teacher, this provides an *enormous* educational benefit: even a full classroom-sized stack of 30 quizzes can (I can attest from experience) be completely graded, and handed back to the students, in 5 minutes or less. This lets the teacher take advantage of one of the other phenomena discussed in Chapter 2: when an incorrect answer is *immediately* corrected, it actually reinforces the *correct* answer.

In other words, the teacher should collect quizzes, grade them right then and there as they come in, hand them back as soon as the process is complete, and then review the questions and answers. Students should be required to write the correct answer next to any question they missed, to further solidify the proper mental connection. Assuming there is not an unusual amount of confusion, a ten-question quiz ought to take between 2 and 5 minutes to review and discuss. Following the discussion, the teacher can re-collect the quizzes for grade book entry purposes.

(Note that it is not actually important to the memory-building goal that the quizzes be worth "real points"—I would suggest doing so only to provide a little extra reason for students to take the enterprise seriously, since I'm certain there will, sadly, be some who are somehow *not* intrinsically motivated by the chance to memorize an interesting and important set of vocabulary words. Kids these days…)

Recording Quiz Data

Whether or not the scores are entered in a grade book, it is essential that the teacher record question-by-question data on student success. I recommend a spreadsheet with vocabulary words along one axis and assessment dates along the other (I have provided an example in Appendix A). What is specifically important to keep records of is the percentage (or at least raw number) of your students who missed each term, so that you know which need re-teaching or re-testing and which are generally solid. Having the second axis will allow you to see progress over time – with luck, the numbers for each term will decrease over time.

There's no *especially* quick way to compile this data, but it's also significantly less urgent to take care of than the initial grading pass. If you've done something handy like mark wrong answers with a slash through the question, then you can lay the quizzes out side by side and just count slashes in each row.

(One final, and minor suggestion for spreadsheet ease-of-use: if you're conversant in Excel or equivalent software, you can set Conditional Formatting rules to turn a cell green, yellow or red based on the appropriate thresholds. I tend to have between 15% and 25% marked yellow, and anything above 25% marked red, so I can at a glance see both

which terms are least well mastered and rough progress in mastering a single term over time.)

Quiz Design

Counterintuitive though it may be to discuss this last, one of the cornerstones of this curriculum is that the design of each quiz is responsive to the class's past performance. With the exception of the very first quiz, which can be assembled literally at random from your pool of terminology, the creation of each subsequent assessment should be seen as the *final* step of processing the previous one.

Specifically, for a 10-question quiz, given Brown, Roediger & McDaniel's research in *Make it Stick* on interleaving, I would suggest the following balance of familiar and unfamiliar terminology:

- 3-4 questions about previously-untested terms (if such exist)
- 4-6 questions re-testing terms which a significant percentage of students missed on a prior assessment
- 1-2 questions re-testing terms which were *not* missed by many students, but which have not reappeared on a quiz for a significant length of time.

A well-kept spreadsheet is absolutely essential for designing assessments in this model, as it provides all the important information immediately and visually. In fact, with a spreadsheet and a little bit of practice, it is possible to select a set of 10 words in something like 2-3 minutes, bringing the total time from "no quiz" to "quiz is fully assembled" into the range of 4-5 minutes.

CHAPTER FIVE

Conclusions

Overview

In full frankness, one of the most important self-reflective lessons I have learned from the process of creating this capstone is that, for all I am a curious and diligent *learner* I am, ironically, not a very good *student*. This project grew out of a task I was set at my school to teach sophomores an extraordinarily lengthy list of exceptionally difficult vocabulary words, and wondering whether there was a way to optimize the process for them or myself – how best to teach disciplinary vocabulary in the secondary classroom. For all that I enjoyed diving into the literature, reading fascinating articles about psychology and language, and pulling ideas from them to test out in my classroom, I found the process of actually writing it all down – turning that work into the actual document you've just read – to be tedious if not actively unpleasant. I'm very proud of the final curriculum I've designed, and also extremely certain that I'm better suited for teaching in high school than higher academia.

In this final chapter, I will be reflecting in much greater detail on the different elements of my capstone project. I will revisit my research and literature review, my curriculum creation process, and the curriculum itself, considering its limitations and possibilities.

The Capstone Process

I took the two pre-capstone classes before I had a clear idea of what I wanted to do with the project, so my initial attempts were quite scattered. The final plan, as I've

mentioned, crystallized when I started working full-time, and had an actual, practical classroom problem to solve. Of course, I needed to start teaching the material immediately, and so the curriculum I actually implemented in my classroom evolved continually over the research and writing process. Originally I wanted to write the capstone as a research project, tracking students' scores in detail over time, but discovered that this slipshod approach to implementation would not play nicely with the IRB's timetable and expectation. Although by my third year teaching the material my approach had mostly stabilized to the curriculum presented in this paper, at that point it seemed simpler to rework the capstone into a curriculum design model – I did not anticipate that it would take me well over another two years to finish the whole thing.

Which brings me to the other important lesson I learned from the process: ADHD is no joke. I was diagnosed in elementary school, but haven't used any sort of formal treatment or medication since at least high school, and as an adult considered it a largely unimportant element of my life. Working on the capstone, however – or, rather, repeatedly failing to – I eventually came to understand that my issues with procrastination and avoidance were, if not *caused* by my condition, unquestionably *exacerbated* by it. So I relearned coping strategies, rebuilt support systems, and recruited my wife to hold me accountable when my own brain declined to do so. While my long-suffering advisor can attest that all these mechanisms hardly cured my fundamental issues with inexplicable delays, they not only helped me finish the capstone at all, but helped me learn how to manage large projects in general.

On a more positive (if equally unsurprising) note, I discovered that for me, genuinely the biggest roadblock was always just *starting* to write. Once I broke the seal,

got a few words down on the page, actually opened up the document to read the comments, etc., the rest of it would flow smoothly and easily. Even the hulking monolith of the literature review, once I got all the materials organized and the rough outline set, took only a week or so to get onto the page. Since I profess to be an English teacher, that's probably a good sign.

The Literature Review

One of the early discoveries that actually reinforced my interest in my research question was that it genuinely had not been explored specifically before. To illustrate the breadth of the gap, the term “disciplinary vocabulary” itself is one I settled on informally, as I was unable to find a standardized term in the literature. In fact, I was literally unable to find a single academic paper via searches for keywords like “disciplinary vocabulary” “technical vocabulary” or several other equivalents – it really seems as if nobody has conducted the particular study I was interested in.

The upside of this absence of directly relevant material was that it turned my literature review process from a straightforward scavenger hunt into something of a puzzle. I had a vague idea of what I was hoping my final curriculum would look like, in terms of its focus on memorization and its long-term goals, and so I used the literature to “triangulate” in on that target. I wanted students to master vocabulary, so I found papers on vocabulary acquisition in the context of second language learning. My school’s default preference was for flashcards, so I looked up the science on flashcard use, hoping to either codify, revise or entirely reject that approach. I wanted students to memorize terms in advance of higher-level applications, so I found my way into the world of non-

educational psychology, where I discovered several researchers investigating the underlying mechanics of memory.

Throughout the full process, the two most useful habits I cultivated were trawling papers' bibliographies and annotating my own. Because so many of the papers I was looking through were themselves only tangentially related to my guiding question, I knew I would have to churn through a ton of them before assembling my core set, so for every paper even marginally relevant – whether or not I added it to my final list – I made a point of reading through the *entire* bibliography, and copy-pasting any titles that sounded potentially interesting into a gigantic spreadsheet I kept open on my desktop. Over the course of the full research process, I must have added at least 125-150 articles to that list before ultimately boiling them down to the ~50 which I discussed in my review chapter.

Moreover, careful annotation of those articles was an absolutely critical choice. I generally learn very well from reading, and don't often need to take notes, but I knew that the sheer scope of this project would likely outpace my own memory, and so I forced myself to write a full paragraph about every single article that I thought I could use, immediately after I read it. These annotations not only saved me when the several months between reading and writing had eroded my memory a bit, but were essential in helping me organize and outline the review chapter itself: I had printouts of every single one on my desk as I worked.

Ultimately, the literature review process was hugely valuable less because it taught me anything *new* about research skills, but because I had never before taken on a project so large that I actually needed to *employ* all the skills I had previously learned. It

seemed greatly intimidating before I started, but having gone through the process I feel much more prepared for any future projects I might need to take on.

The Final Project

The curriculum I have designed is, based on my review of several other curriculum-design capstones in Hamline's archive, unusual in a few ways. First, it is **subject-agnostic** – while I provide some examples of how I might implement it in a Rhetoric classroom, there is nothing about the processes themselves tied to that specific field. Outside of English, I can see the curriculum being extremely useful in many science classrooms (I know that freshman biology at my school, for instance, involves well over a hundred vocab words), history, or even math. It's possible that it may be useful in a foreign language class, but there are differences between learning new words in a primary language vs. as part of acquiring a second one and I'm not confident that my approach would be *optimal* in the latter circumstance.

Second, my curriculum is **modular**, designed to fit into the middle of other, larger units. My recommended time commitment is ~5-10 minutes, 2-3 times a week, which means that teachers could, for example, treat it like a warm-up exercise, rather than a stand-alone unit that's the primary driver of several weeks of class time. That said, I know how precious even 5-10 minutes of class time can be, which is why the curriculum design document also discusses the rationale behind each element. My hope is that teachers considering using the curriculum can, via understanding the intended function of each component, tweak the proportions to fit their needs without sacrificing the underlying utility.

Limitations and Avenues for Further Exploration

Other than the caveat mentioned above regarding second language acquisition, I see two primary limitations of this curriculum. The first is that the entire exercise is predicated on the belief that there is value in rote memorization as a precursor to deeper learning. I offered a non-research-based defense of this in my literature review chapter, but ultimately this is a question more about teaching philosophy than anything else. Many teachers and districts prefer approaches where term acquisition is part and parcel of the broader lesson plans, rather than a separate initial step, and as there is not (as far as I could find) any research directly comparing outcomes for these different approaches, for such classrooms this curriculum would simply not be an appropriate fit.

The other major limitation is that my study did not engage deeply with the potential complications of different learning styles or disabilities. I do feel that most of the literature I pulled from to create the plan is concerned with the mechanisms of memory on such a "deep" psychological level that the implications are *likely* to be largely universal, but it must be acknowledged that I did not seek out research on (e.g.) memory differences in learning-disabled students. To be honest, that seemed like a monumental addition to an already fairly massive project.

It would, however, be one important area for potential further research. I am not a specialist in Special Education, but I've learned enough from my education classes to understand just how vast that field of study is. I do not feel remotely qualified to adapt this curriculum to individualized student needs with the same research-based rigor that went into its initial development, but I'm optimistic that a scholar who already has a solid grounding in the relevant theories and research could do so with much less trouble.

Even without further adapting the curriculum, however, I feel that there is potentially important work to be done in *testing* it in a formal study. As I've mentioned, to the best of my knowledge nobody has asked my specific research question before, and while I've devised an answer that *ought* to work, and which *seemed* like it was working in my informal classroom implementation, that's not the same as a controlled experiment. I strongly suspect that I myself will not be the one to conduct such a test – the length of time it's taken me to complete this Capstone suggests that I'm not especially well-suited for academic publishing – but I hope that someone either tries out my curriculum or develops their own answer to the question, because it remains an important and oddly unanswered one in modern education.

Final Thoughts

Looking back on the whole process, I learned much more from writing my Capstone than I at first cynically assumed I would. Part of what I've learned, of course, is that I really, *really* should not pursue a career in higher academia; fortunately, my classroom teaching work over the same period has reinforced how much I love the career I *have* chosen. However, the practical experience I gained with the research process, forcing myself through it, prepared me better than any abstract understanding for teaching my own students about how to pursue their own research – and the scale of my literature review was a fantastic tool for squashing their complaints about having to find *five whole sources*. Moreover, what I learned from my research itself – the depth of knowledge I gained about the mechanics of memorization, testing and learning – has both deepened my own teaching theory and practice and been a consistent source of interest for my colleagues. I have taught at two separate schools, two thousand miles apart, since

beginning this project, but the deep knowledge I have gained through my work has been a direct benefit to the teachers and students at both. Ultimately, while I do regret the time I've wasted *not* working actively on my Capstone over the past four years, I don't regret for a moment the hours I *did* put into it, and I look forward to continuing to reap the benefits of my efforts for years to come.

REFERENCES

Albers, C. A., & Hoffman, A. (2012). Using flashcard drill methods and self-graphing procedures to improve the reading performance of English language learners. *Journal of Applied School Psychology, 28*(4), 367-388. doi:10.1080/15377903.2012.731365

Blachowicz, C. L. Z., Fisher, P. J. L., Ogle, D., & Watts-Taffe, S. (2006). Theory and research into practice: Vocabulary: Questions from the classroom. *Reading Research Quarterly, 41*(4), 524-539. Retrieved from http://www.jstor.org/stable/4151816

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan, 80*(2), 139-144, 146-148. Retrieved from http://www.jstor.org/stable/20439383

Bretz, S. L. (2001). Novak's theory of education: Human constructivism and meaningful learning. *Journal of Chemical Education, 78*(8), 1107. doi:10.1021/ed078p1107.6

Brown, B. A., & Ryoo, K. (2008). Teaching science as a language: A "content-first" approach to science teaching. *Journal of Research in Science Teaching, 45*(5), 529-553. doi:10.1002/tea.20255

Brown, T. S., & Perry, F. L.,Jr. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly, 25*(4), 655-670. Retrieved from http://www.jstor.org/stable/3587081

Butler, A., & Roediger, H. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4-5), 514-527. doi:10.1080/09541440701326097

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*(6), 633-642. doi:10.3758/BF03202713

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380. doi:10.1037/0033-2909.132.3.354

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2)

Craik, F. I. M., & Tulving, E. (1975). *Depth of processing and the retention of words in episodic memory.* Journal of Experimental Psychology: General. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ128798&site=ehost-live

Debb, S. M., & Debb, S. M. (2012). Teaching introductory psychology in the community college classroom: Enhancing student understanding and retention of essential information. *Inquiry, 17*(1), 27-36.

Don L. F. Nilsen, & Nilsen, A. P. (2009). Naming tropes and schemes in J. K. rowling's harry potter books. *The English Journal, 98*(6), 60-68. Retrieved from http://www.jstor.org/stable/40503461

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition; Stud.Sec.Lang.Acq., 21*(2), 285-301.

Fata-Hartley, C. (2011). Resisting rote: The importance of active learning for all course learning objectives. *Journal of College Science Teaching*, 40(3), 36-39.

Fisher, D., Grant, M., & Frey, N. (2009). Science literacy is > strategies. *Clearing House, 82*(4), 183-186. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=keh&AN=36593454&site=ehost-live

Freedman, J. S. (1986). Cicero in sixteenth- and seventeenth-century rhetoric instruction. *Rhetorica: A Journal of the History of Rhetoric, 4*(3), 227-254. Retrieved from http://www.jstor.org/stable/10.1525/rh.1986.4.3.227

Fritz, P. A., & Weaver, R. L. (1986). Teaching critical thinking skills in the basic speaking course: A liberal arts perspective. *Communication Education, 35*(2), 174. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=9452319&site=ehost-live

Gelder, T. v. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching, 53*(1), 41-46. Retrieved from http://www.jstor.org/stable/27559216

Grillo, K. J., & Dieker, L. A. (2013). A new twist on vocabulary instruction for students with learning disabilities in biology. *The American Biology Teacher, 75*(4), 264-267. Retrieved from http://www.jstor.org/stable/10.1525/abt.2013.75.4.7

Grove, N. P., & Lowery Bretz, S. A continuum of learning: From rote memorization to meaningful learning in organic chemistry. *Chemistry Education Research and Practice,* (3), 201. doi:10.1039/C1RP90069B

Haynie,W.J.,,III. (1997). Effects of anticipation of tests on delayed retention learning. *Journal of Technology Education, 9*(1), 20-30.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290-296. doi:10.1037/a0028468

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-58. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ632733&site=ehost-live

Kail, S. R. (2008). Vocabulary instruction goes "old school". *The English Journal, 97*(4), 62-67. Retrieved from http://www.jstor.org/stable/30047249

Khoii, R., & Sharififar, S. (2013). Memorization versus semantic mapping in L2 vocabulary acquisition. *ELT Journal, 67*(2), 199-209. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ998025&site=ehost-live; http://dx.doi.org/10.1093/elt/ccs101

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*(9), 1297-1317. doi:10.1002/acp.1537

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*(6), 585-592. Retrieved from http://www.jstor.org/stable/40064958

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*(4), 449-468. doi:10.1037/a0017350

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied, 19*(1), 1-13. doi:10.1037/a0032147

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989-998. doi:10.1037/a0015729

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=44400936&site=ehost-live

Lessard-Clouston, M. (2008). Strategies and success in technical vocabulary learning: Students' approaches in one academic context. *Indian Journal of Applied Linguistics, 34*(1-2), 31-63. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ894238&site=ehost-live; http://bahripublications.in/appliedlinguitic/back_volume.htm

MacQuarrie, L. L., Tucker, J. A., Burns, M. K., & Hartman, B. (2002). Comparison of retention rates using traditional, drill sandwich, and incremental rehearsal flash card methods. *School Psychology Review, 31*(4), 584-95. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ667629&site=ehost-live

Martin-Chang, S. L., Levy, B. A., & O'Neil, S. (2007). Word acquisition, retention, and transfer: Findings from contextual and isolated word training. *Journal of*

Experimental Child Psychology, 96(1), 37-56. doi:http://dx.doi.org.ezp1.lib.umn.edu/10.1016/j.jecp.2006.08.004

Mattern, K. D., Shaw, E. J., Xiong, X., & Board, C. (2009). *The relationship between AP[R] exam performance and college outcomes. research report no. 2009-4*College Board. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED521458&site=ehost-live; http://professionals.collegeboard.com/profdownload/pdf/09b_269_RReport2009_Groups_Outcomes_WEB_091124.pdf

Mercer, S. (2005). Vocabulary strategy work for advanced learners of English. *English Teaching Forum, 43*(2), 24.

Miller, J. B. (2005). Practicing the ancient art of "memoria" in the modern classroom. *Communication Teacher, 19*(2), 48.

Morin, R., & Goebel, J. (2001). Basic vocabulary instruction: Teaching strategies or teaching words? *Foreign Language Annals, 34*(1), 8-17. doi:10.1111/j.1944-9720.2001.tb02797.x

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3-8. doi:10.1037/0278-7393.31.1.3

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*(3), 243-257. doi:10.1037/a0016496

Roediger III, Henry L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255. Retrieved from http://www.jstor.org/stable/40064526

Roediger III, Henry L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210. Retrieved from http://www.jstor.org/stable/40212166

Roediger, I., Henry L. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology, 59*(1), 225-254. doi:10.1146/annurev.psych.57.102904.190139

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99-115. doi:http://dx.doi.org.ezp3.lib.umn.edu/10.1016/j.jml.2014.03.003

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72-110. Retrieved from http://www.jstor.org/stable/1170287

Storm, B., Bjork, R., & Storm, J. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38*(2), 244-253. doi:10.3758/MC.38.2.244

Symonds, P. M., & Chase, D. H. (1992). Practice vs. motivation. *Journal of Educational Psychology, 84*(3), 282-289. doi:10.1037/0022-0663.84.3.282

Volpe, R., Mulé, C., Briesch, A., Joseph, L., & Burns, M. (2011). A comparison of two flashcard drill methods targeting word recognition. *Journal of Behavioral Education, 20*(2), 117-137. doi:10.1007/s10864-011-9124-y

Westheimer, J. (2010). No child left thinking: Democracy at risk in canada's schools. *Education Canada, 50*(2), 5-8. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=eue&AN=508156825&site=ehost-live

Wetzels, S. A. J., Kester, L., van Merriënboer, J. J. G., & Broers, N. J. (2011). The influence of prior knowledge on the retrieval-directed function of note taking in prior knowledge activation. *British Journal of Educational Psychology, 81*(2), 274-291. doi:10.1348/000709910X517425

Yager, R. E. (1983). The importance of terminology in teaching K-12 science. *Journal of Research in Science Teaching, 20*(6), 577-588. doi:10.1002/tea.3660200610

APPENDIX A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		1A	2A	3A	4A	5A	6A	7A	8A	9A	10A	11A	12A	13A	14A
2		8.04	8.78	7.81	8.00	8.37	7.65	7.59	8.56	8.33	8.15	7.44	7.11	6.74	8.21
3	Definition		2												
4	Genus	12		6	6				3						
5	Division (T)	7		7	4							6			1
6	Similarity		2												
7	Difference	7		2											
8	Degree	10		4						4					
9	Cause and Effect		6						4						
10	Antecedent & Consequence		13	13	4					6				11	2
11	Contraries	5				11		7					8		
12	Contradictions		5				5				6			7	
13	Possible and Impossible	3													
14	Past Fact and Future Fact		0												
15	Authority	1													
16	Testimonial		2												
17	Statistics		0												
18	Maxims		0												
19	Law	0													
20	Precedent		3								2				
21	The Worthy	4				1									
22	The Advantageous	4					3								
23															
24	Declarative							9		0					
25	Imperative			4											
26	Interrogative					0									
27	Exclamatory				12			5			3				
28	Simple						0								
29	Compound			8	1										
30	Complex						13		5					7	
31	Compound-Complex					5						3			
32	Loose				7		5					5			
33	Periodic					3									
34	Natural order				6					6			2		
35	Inverted order			6							0				
36															
37	Abstract			4				2							
38	Concrete					4							4		
39	Formal						0								
40	Informal				3							6			
41	Common				6		3								
42	Jargon					1									
43	Referential				5				0						
44	Emotive			3											

A sample tracking spreadsheet, as described in Chapter 4. Each column represents a quiz date; numbers in cells are a count of students (out of 27, in this case) who failed to identify the term.