

Effectiveness and Adoption of a Drawing-to-Learn Study Tool for Recall and Problem Solving: Minute Sketches with Folded Lists

Paul D. Heideman,* K. Adryan Flores, Lu M. Sevier, and Kelsey E. Trouton

Department of Biology, College of William and Mary, Williamsburg, VA 23187

ABSTRACT

Drawing by learners can be an effective way to develop memory and generate visual models for higher-order skills in biology, but students are often reluctant to adopt drawing as a study method. We designed a nonclassroom intervention that instructed introductory biology college students in a drawing method, minute sketches in folded lists (MSFL), and allowed them to self-assess their recall and problem solving, first in a simple recall task involving non-European alphabets and later using unfamiliar biology content. In two preliminary *ex situ* experiments, students had greater recall on the simple learning task, non-European alphabets with associated phonetic sounds, using MSFL in comparison with a preferred method, visual review (VR). In the intervention, students studying using MSFL and VR had ~50–80% greater recall of content studied with MSFL and, in a subset of trials, better performance on problem-solving tasks on biology content. Eight months after beginning the intervention, participants had shifted self-reported use of drawing from 2% to 20% of study time. For a small subset of participants, MSFL had become a preferred study method, and 70% of participants reported continued use of MSFL. This brief, low-cost intervention resulted in enduring changes in study behavior.

INTRODUCTION

Visual representations are ubiquitous in science as a tool for teaching, understanding, communicating, and developing ideas (Van Meter and Garner, 2005; Quillin and Thomas, 2015). Visual representations in primary research publications are used routinely to present new hypotheses, and many of these visual representations eventually develop into abstract visual models presented to students as illustrations in textbooks. Visual representations of concepts drawn by students can help them recall (Wammes *et al.*, 2016), think, generate hypotheses, develop predictions and experiments, and communicate results (Van Meter and Garner, 2005; Schwarz *et al.*, 2009; Ainsworth *et al.*, 2011; Quillin and Thomas, 2015). Drawing may be most useful when it actively engages students in selecting, organizing, and integrating information to develop a visual model that represents a mental model (Van Meter and Garner, 2005; Mayer, 2009; Quillin and Thomas, 2015). Learner-generated visual models drawn by students can aid in their acquisition of knowledge and their communication of ideas to others (Van Meter and Garner, 2005) while serving as an important tool to aid in problem solving (Quillin and Thomas, 2015). The literature cited earlier suggests that drawing may aid students in learning tasks from the simplest, such as developing memory for core content, to the most complex, including hypothesis generation, prediction, and analysis.

Instructors in biology commonly draw models in aid of understanding or when presenting problems to students, but generally not with an explicit goal of developing drawing skills in students for use in thinking and modeling (Quillin and Thomas,

Deborah Allen, *Monitoring Editor*

Submitted June 6, 2016; Revised March 8, 2017; Accepted March 21, 2017

CBE Life Sci Educ June 1, 2017 16:ar28

DOI:10.1187/cbe.16-03-0116

*Address correspondence to: Paul D. Heideman (pdheid@wm.edu).

© 2017 P. D. Heideman *et al.* CBE—Life Sciences Education © 2017 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

2015). It may be important to develop interventions that build student skills (Van Meter and Garner, 2005; Leutner *et al.*, 2009; Schwarz *et al.*, 2009; Quillin and Thomas, 2015), because, for some tasks, drawing has not produced gains in learning (see discussion in Leutner *et al.*, 2009; Quillin and Thomas, 2015). While drawing-to-learn shows promise as a learning tool for students, an important challenge is the development of interventions that successfully motivate students to draw and improve skills in drawing and model-based reasoning (Van Meter and Garner, 2005; Quillin and Thomas, 2015).

It is challenging to motivate college students to change study methods, even though students commonly use study methods that are known to be ineffective compared with more active study methods (Fu and Gray, 2004; Dunlosky *et al.*, 2013). Students hesitate to adopt new learning methods, even when they have evidence for an effective replacement (Pressley *et al.*, 1989; National Research Council [NRC], 2001; Fu and Gray, 2004). This reluctance to apply a new method is understandable when students lack experience and metacognitive skills to apply new methods and self-monitor the effectiveness of their learning (Bielaczyc *et al.*, 1995; Bransford, 2000; Tanner, 2012). Students may require both metacognitive skills to assess new study methods (Schraw *et al.*, 2006) and personal experience with an effective new study method. In our experience teaching drawing to college students for model-based reasoning, the challenge has been in demonstrating short-term gains in return for their efforts to produce and practice drawing. Students need practice and experience with explicit instruction on the value of investing time and effort on learner-generated drawings. Our informal interventions with individual students over many years is consistent with proposals from others: an effective intervention requires 1) teaching an effective method for drawing (Van Meter and Garner, 2005), 2) practice sessions with feedback (Schwarz *et al.*, 2009; Quillin and Thomas, 2015), 3) an explanation of drawing for self-regulated learning as a metacognitive strategy (Bransford, 2000; NRC, 2003), and 4) having each individual self-assess the utility of drawing for learning gains he or she values.

Here, we describe results of an intervention with three goals for students: 1) learning a drawing method offering greater effectiveness of study for recall, understanding, and/or problem solving than the passive study methods preferred by most of our students; 2) self-scoring assessments to self-evaluate effectiveness; and 3) adoption by students of the method, if effective, for enduring, routine use. This drawing-to-learn intervention is consistent with research on learning and applications of drawing methods in teaching (see Van Meter and Garner, 2005; Forbus *et al.*, 2011; Jee *et al.*, 2014; Wammes *et al.*, 2016) but was largely developed empirically, in an iterative process involving many individual students over many years. The drawing strategy uses a minimalist approach to drawings (termed “minute sketches,” see *Methods* for more details), potentially minimizing the time necessary for practice and minimizing cognitive load (Sweller, 1988; Sweller and Chandler, 1994; Schnotz and Kürschner, 2007; Leutner *et al.*, 2009; Verhoeven *et al.*, 2009). In this approach, minute sketches are associated with terms using retrieval practice (Karpicke and Blunt, 2011) with a “foldable” or “folded list” (see *Methods*). In combination, minute sketches with folded lists (MSFL) are intended to be a study method to

develop recall and understanding, producing learning outcomes that students value sufficiently to adopt drawing as a long-term study method. Minute sketches are intended for active-learning applications (Freeman *et al.*, 2014) in higher-order cognitive processing, in which students learn to view each minute sketch as a hypothesis for a structure or a process, with the ability to manipulate the visual model to develop predictions and solve problems. Our ultimate goals for students when using minute sketches includes hypothesis generation, making predictions, and problem solving, but students in science, technology, engineering, and mathematics (STEM) also need to have skills at recalling extensive amounts of discipline-based content. It may be the latter that motivates students to adopt drawing-to-learn.

The study was both an intervention and a quasi-experimental test of MSFL as a study tool. In three experiments, we tested whether students performed better in recall and/or problem solving with MSFL than with a preferred study method: visual review (VR; see *Results* for evidence that, for our students, VR is a preferred study method). In the third experiment, we also assessed whether an intervention might result in increases in drawing as a study method. We used pre-post surveys to assess the use of drawing for study in the intervention group and a comparison group. On the basis of the results, we discuss potential uses of MSFL to help students learn, practice, and self-assess this addition to their study methods.

METHODS

Minute Sketches

A minute sketch (Figure 1) is a simple diagram, ideally user generated (Quillin and Thomas, 2015), that uses the minimum number of lines and symbols necessary to represent a concept fully to a user, making it reproducible in less than a minute. The 60-second limit prevents wasteful overelaboration, enforces focus on essentials, and reduces risk of cognitive overload (Sweller, 1988; and see coherence principles 1 and 3 in Mayer, 2009). All terms needed with the sketch are placed in a list to one side (an application of multimedia learning; Mayer, 2009), never on the sketch itself, because humans can find simultaneous processing of images and text challenging (Figure 1). Sketches may include numbers to indicate sequence or letters that are commonly interpreted as symbols (e.g., Na⁺ for sodium ions or H₂O for water). The components of the sketch should always remind the user of any necessary explanation; a written explanation is not allowed. Explanations in multimedia learning may cause an unhelpful “redundancy effect” (Sweller and Chandler, 1994; Mayer, 2009), in which redundant text added to a diagram increases cognitive load and inhibits learning (Sweller and Chandler, 1994). If a sketch is sufficient to trigger recall and understanding, then redundant text wastes time and energy to process that unnecessary content (Schnotz and Kürschner, 2007).

Any minute sketch is also a hypothesis about structure, function, or relationships. Minute sketches typically include relationships broadly comparable to those in structure–behavior–function (SBF) models that connect structure to behavior and function in biology (Dauer *et al.*, 2013; Speth *et al.*, 2014). Changes may be made to the original minute sketch, similar to alternate pathways in an SBF model (e.g., see Figure 1 in Speth *et al.*, 2014), allowing the newly manipulated

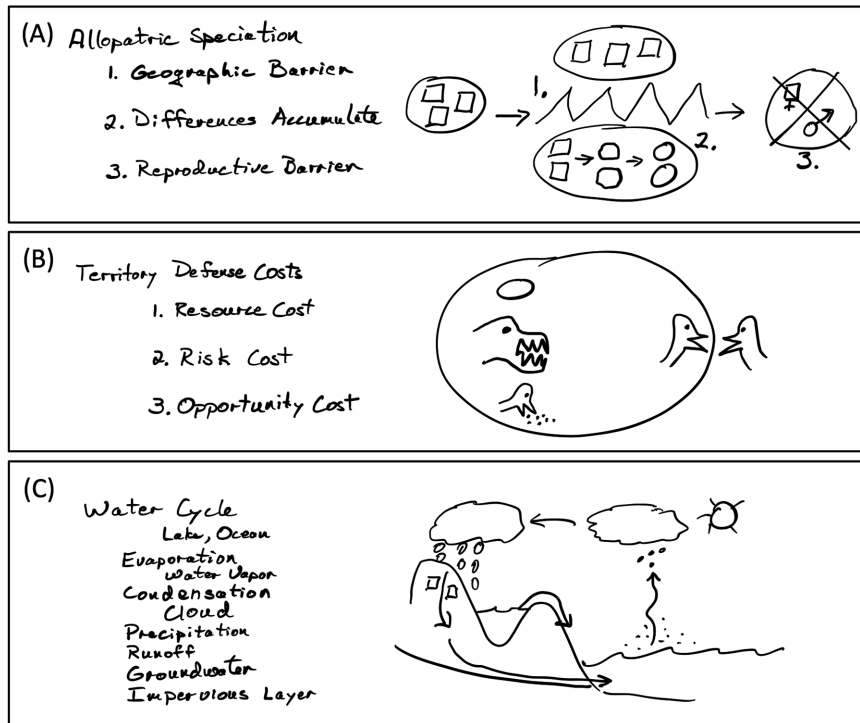


FIGURE 1. Minute sketches for three concepts, each instructor drawn based on student examples. (A) Allopatric speciation is hypothesized to occur when individuals (squares at left) in an original single population (circle at left) are isolated by a geographic barrier such as an ocean or a mountain range (jagged line at center labeled “1”). Over time (arrows in lower oval) one population changes (squares becoming circles in the lower oval), while the other may or may not change (upper oval with squares). Eventually, even if individuals encounter each other, they have accumulated too many differences for successful reproduction (circle with X labeled “3”). (B) In a hypothesis for the costs of territory defense, there are three costs of defending a territory, indicated by the boundary line, from others of the same species: birds on the right side facing each other across the territory boundary. Resources used for territory defense cannot be used for other purposes, such as reproduction (the egg, representing resource cost). Vigilance and fighting for territory defense increases risks, such as risk of predation (the predator head, representing risk cost). The time spent defending the territory is an opportunity cost, time lost from alternative activities, such as feeding (the bird head pecking at food, representing opportunity cost). In B, note that the term for each cost has a location that corresponds to its representation in the sketch. (C) The water cycle. Each of the three minute sketches represents the essentials of the concept or hypothesis, and elements in the sketches may be manipulated to allow the user to predict outcomes (see the text).

sketch to be used in predicting changes to the system. For example, in the minute sketch for the water cycle (Figure 1C), a student might choose to increase atmospheric dust particles. Using the sketch, a student can identify potential changes through each succeeding step in the entire water cycle (e.g., less sunlight reaching water, causing less evaporation, causing ...). As students use minute sketches for more complex concepts, earlier minute sketches appear in simplified form, consistent with findings that representations may become simpler with increasing expertise (Hay *et al.*, 2013; Dowd *et al.*, 2015). For example, in the minute sketch for the water cycle (Figure 1C), symbols exist for evaporation (dots rising from the ocean) and condensation (dots becoming small droplets), each of which can be elaborated further in separate minute sketches,

producing an interconnected web. Students are capable of misapplying or misunderstanding symbols such as arrows or connections within sketches or diagrams (Novick and Catley, 2007; Wright *et al.*, 2014), and so an important step in developing minute sketches is checking understanding by referring back to instructional materials or instructors.

Folded Lists

In addition to understanding, learning requires fluent recall of essential concepts with associated terms. The tool for fluent recall of a minute sketch with associated terms is a folded list or “foldable” (Figure 2) for retrieval practice. Students view only one column at a time, either minute sketches or terms, hiding the other column. The hidden column is rewritten or resketched from memory, checking a hidden column when needed (Figure 2). Students are encouraged to recheck the hidden column when needed in order to minimize a potential negative effect of guessing (Roediger and Marsh, 2005; McDermott, 2006; Chang *et al.*, 2010). When sketching, students are asked to think about each term as they sketch, referring back to instructional materials as needed to check their understanding. When writing terms, students are asked to think about the connection between each term and the sketch. Students may vocalize terms and explain sketches aloud as they draw (see the modality principle in Mayer, 2009). Students may save time in later iterations by abbreviating or removing terms or lines they recall easily (Figure 2, panels on right side). When students can readily reproduce and explain the sketch from the terms and the terms from the sketch, they know they have achieved fluent recall with their current understanding

on that day. If they cannot, students have an unambiguous self-diagnosis: they need more practice.

Experiments

Informal feedback from students suggests that after practicing two complete cycles (sketch to terms and terms to sketch repeated twice) on three different days, students have fluent recall with their current understanding for at least 1 day, and usually for more. Similar feedback suggests that, weeks or months later, before a final exam or when reviewing content for advanced courses, practice on as little as one additional day recovers their recall and understanding. This feedback was incorporated into the design of the experiments discussed here.

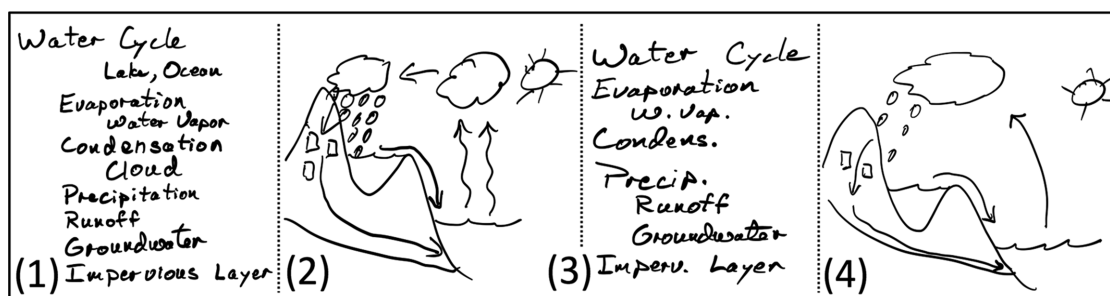


FIGURE 2. Representation of a minute sketch and list of terms practiced in a folded list in a series of steps (example was drawn by an investigator based on many student examples). 1) The user writes the list of key terms. 2) While thinking through the terms, the user redraws the minute sketch from memory. 3) After folding the terms under the paper (or covering the terms), the user rewrites the terms from memory while thinking through the minute sketch. 4) After hiding the sketch, the user redraws the sketch from memory while thinking through the terms. As users gain familiarity with the terms and sketch, they are encouraged to skip obvious terms (e.g., lake, ocean, cloud), abbreviate when the abbreviation is a sufficient reminder (e.g., “w. vap.” and “precip.”), and reduce the sketch to essentials. In the sketch on the right, there is one cloud instead of two, fewer lines of raindrops, and fewer arrows. Any line that fails to add meaning can be eliminated. This example shows a typical simplifying progression.

The first two experiments were intended to assess the potential of MSFL for a simple learning task. Experiment 1 was intended to assess MSFL in comparison to VR when learning simple content in the absence of information on learning and study methods. This experiment compared students using either MSFL or VR on a simple non-STEM learning task: learning characters from non-European alphabets with associated phonetic spellings of sounds. While the goal of the experiment was not explicitly stated, information to participants was presented in such a way that the experiment could be interpreted reasonably as a test of the difficulty of learning different alphabets. A second experiment (experiment 2) gave each participant the same learning task, but in a crossover design with matched study time on equivalent content, with one page studied by MSFL and the other by VR.

The third experiment (experiment 3) was an intervention in which participants studied using MSFL in comparison with VR on matched content with self-assessment of their learning with the two study methods. The first part of experiment 3 allowed students to assess whether their retention of simple content, unfamiliar non-European alphabets, was better using MSFL or VR. Importantly, this part of the experiment also allowed investigators to correct misapplications of MSFL as a study method. The second part of experiment 3 allowed students to self-assess whether their retention and problem solving on unfamiliar biology content was better after studying using MSFL or VR.

Experiment 1. Between-Subjects Test of MSFL and VR. In an ex situ experiment with participants recruited from an organic chemistry course ($N = 33$; provided with a \$15 gift card), we assessed MSFL in comparison with VR when learning simple content in the absence of information on learning and study methods. The non-STEM learning task was to learn characters from non-European alphabets with associated phonetic spellings of sounds, a task similar to learning symbols with pronunciations in STEM courses (such as μ or Å with an associated sound or term, micron and angstrom, respectively). The use of characters from an unfamiliar non-European language (factor: language—Arabic or Korean) allowed us to control for prior knowledge. Participants were assigned a language at random; participants

already familiar with an assigned alphabet were reassigned to the alternative. Participants were asked to learn 12 characters and their phonetic spellings using one of two study methods (factor: study method—MSFL or VR). Separate sessions were held for the VR and MSFL treatment groups. Participants were allowed to infer that the study’s purpose might be to assess the difficulty of learning different alphabets. After a survey on demographics and study methods and an introduction to the character sets, participants were given 5 minutes of instruction on their study method and 3 minutes of practice studying their character set. Participants were instructed to practice on their own for 5 minutes on each of three different days, recording start and end times on each day (see timeline in Figure 3). Practice was not allowed on the day of the quiz. The MSFL group was provided with pages for practice and asked to submit these practice sheets. In a quiz 7–10 days later, participants reproduced their character set with phonetic spellings. Quizzes were scored by the investigators, blind with respect to treatment. Two points were awarded for a correct pairing of character with phonetic sound, only one point for correct characters with no phonetic spelling or a phonetic spelling that was mismatched, and $\frac{1}{2}$ point was removed for minor errors (such as a slightly misdrawn character) according to a rubric (maximum score = 24).

Experiment 2. Within-Subjects Test of MSFL and VR. In a crossover design, this ex situ experiment asked each participant to study one page of content (non-European alphabets) using MSFL, and a second, equivalent page using VR. Participants were recruited from college calculus courses ($N = 23$; provided with a \$20 gift card). Participants were assigned randomly to the Korean or Arabic alphabet (factor: language); those reporting familiarity with an assigned alphabet were reassigned to the other alphabet. Each participant studied two content pages (factor: content page) of equivalent difficulty that were assigned at random for study using MSFL or VR. The session on day 1 followed methods of experiment 1, except that all participants received instruction with both methods, and all practiced one page with VR and one page with MSFL. Practice before the quiz 7–10 days later was as in experiment 1, except that each participant practiced one set of characters with VR and the other set

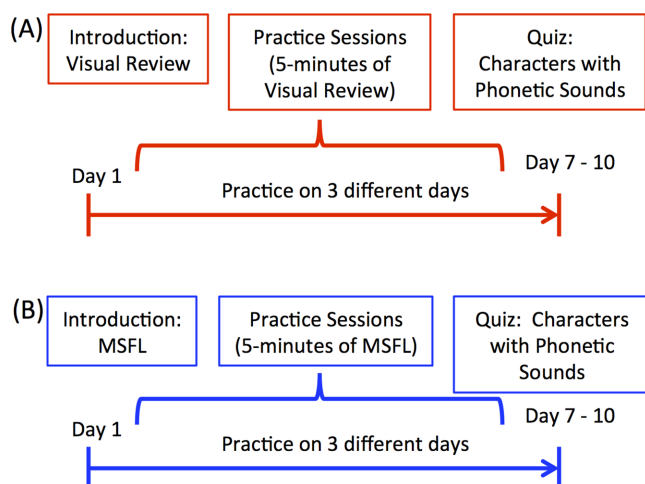


FIGURE 3. Timeline for experiments 1 and 2: (A, in red) VR of 12 characters and associated phonetic spellings practiced for 5 minutes on each of three different days before the day of a quiz, and (B, in blue) practice using minute sketches with folded lists (MSFL) of the same 12 characters and associated phonetic spellings for 5 minutes on each of three different days before the day of a quiz. In experiment 1, the only experiment in which participants used a only a single study method, each participant had only one set of 12 characters and followed the study method in either A or B. In experiment 2, each participant was given two pages of alphabet characters and sounds and studied one page using VR and the other page using MSFL. See *Methods* for additional details.

with MSFL, each for 5 min, on three different days. There was no practice on the day of the quiz. The quiz was administered and scoring was conducted as in experiment 1.

Experiment 3. Within-Subjects Intervention Test for Effectiveness and Adoption of MSFL. This multisession intervention tested whether students who applied the two methods might change their study methods after self-assessment of their learning using MSFL. Participants were recruited from a first-semester biology course (an introduction to ecology, evolution, Mendelian genetics, and biodiversity); participating in the study was one of several methods through which they earned ~1% of the course credit. The intervention began with a survey and a 25-minute workshop on study methods, including evidence that passive VR has low effectiveness (summarized from Dunlosky *et al.*, 2013). Character sets and instructions were presented as in experiment 2. Two participants who were familiar with both Arabic and Korean were assigned character sets from Tamil. The subsequent three sessions included quizzes that were self-scored by participants and later scored by the investigators, blind with respect to study method (for correlations between self-scores and investigator scores, see *Results*). The intent of self-scoring was to make participants aware of any differences in their learning after studying with the two different methods. The pattern of practice sessions and subsequent quizzes was intended to match studying that might be conducted before an exam (quiz 1), followed by a period of time without study following an exam (quiz 2), and later a single study session that might come before a final exam (quiz 3; see timeline in Figure 4).

In session 4, all continuing participants received a new 20-minute presentation on content that was unrelated to any course content (thyroid hormone regulation and metabolism), including two pages of diagrams and accompanying typed text (see the Supplemental Material). The two sets of diagrams were approximately matched for difficulty. During the presentation, brief mention was made that minute sketches might be useful as an aid in solving problems. Next, participants received instructions on minute sketching and were told to generate their own minute sketches based on the content provided. As with the character sets, participants studied one page using VR and the other page using MSFL, assigned at random. Before session 5, participants studied each page for 5 minutes, using VR for one page and MSFL for the other. In session 5, participants were asked to reproduce the content, with explanations, along with two problems to solve using the content. Before session 6, 2 weeks later, participants were told to do no additional studying. Before session 7, held 2 weeks after session 6, participants were asked to study both pages on the night before the session, one page as assigned with VR and the other with MSFL. Again, the pattern of practice sessions was intended to match studying that might be conducted before an exam (session 5), time passing after an exam with no additional study (session 6), and, finally, a single study session that might come before a final exam (session 7; see timeline in Figure 4). In each of the three sessions, quizzes were self-scored by students following a rubric. Quizzes were later scored by the investigators, blind with respect to study method.

Four months after session 7, a subset of participants were recruited from the intervention group (\$35 gift card) for a brief intervention on problem solving using minute sketches. After a presentation on using MS for problem solving, participants received a 25-minute lecture with diagrams on immune function. Participants were randomly assigned one of the two pages to practice for recall and problem solving using VR (10 minutes on each of 3 days), and the other page to practice content and problem solving using MSFL (10 minutes on each of 3 days). One week later, participants solved problems on both pages of content. In addition to the self-scoring by participants, investigators scored all quizzes, blind with respect to study method.

The survey on study methods (pre) administered to participants in the intervention group ($N = 69$) was also issued to a comparison group drawn from the same class ($N = 55$). The same survey was readministered (post) 7 months later (intervention group: $N = 44$ of the 56 participants who completed at least two sessions for the intervention, a 75% participation rate; $N = 57$ for the comparison group from the same course). Each survey asked for percentage of study time allocated to the following study methods: 1) rereading notes, 2) rereading presentations (such as PowerPoint), 3) rereading the textbook, 4) rewriting notes, 5) summarizing (from notes, a presentation, or a textbook), 6) highlighting, 7) flash cards, 8) drawing or sketching, 9) practicing by redrawing, 10) practice testing, 11) self-testing, 12) writing one's own exam questions, 13) mnemonics, 14) "chunking" content, 15) retrieval practice, 16) using folded lists/foldables, and 17) minute sketching. Participants were also asked to indicate the percentage of time in which they used distributed practice. The full survey is presented in the Supplemental Materials. Respondents were given the option to add study methods, but few added any

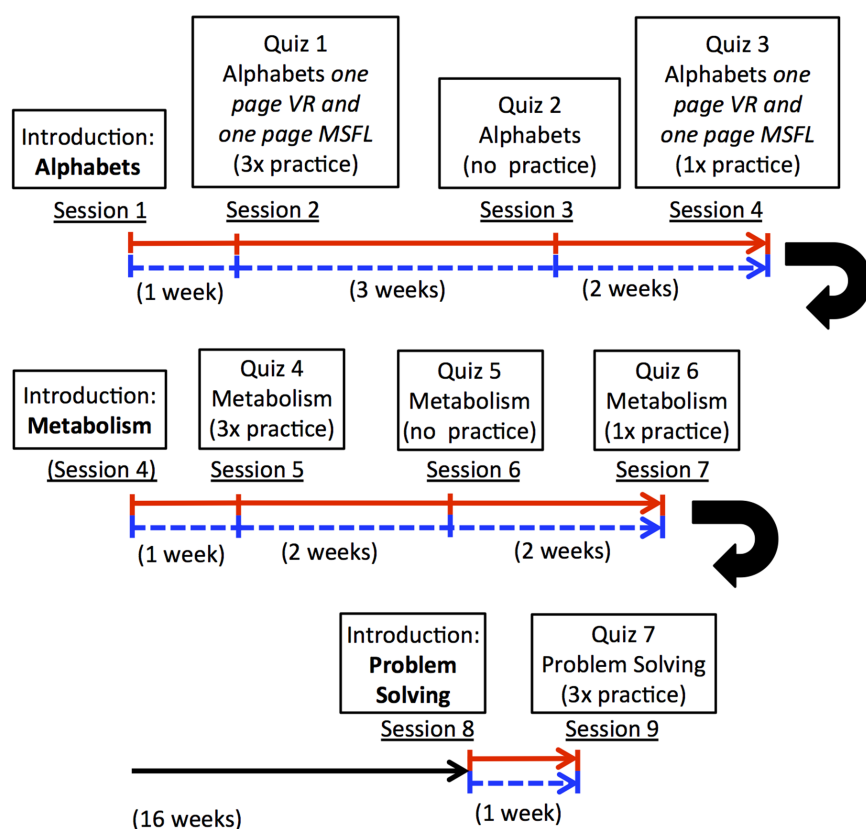


FIGURE 4. Timeline for experiment 3. In experiment 3, the first two sessions were identical to experiment 2, with participants instructed to do time-matched study of alphabet characters and associated phonetic spellings of sounds from non-European alphabets using VR (solid line in red) and MSFL (dashed line in blue) followed by a quiz (session 2). Experiment 3 continued with a second quiz after 3 weeks without study (session 3), and a third quiz after review (session 4). After the quiz in session 4, participants were given novel biology content in two sets of matched pages, one set for study with VR and the other with MSFL, followed by a quiz in session 5, another quiz after 2 weeks without study (session 6), and a final quiz after review (session 7). Sixteen weeks later, a subset of participants received novel biology content (session 8) and additional instructions on using minute sketches to solve problems. Following matched study time on matched content, participants took a quiz that included only problems to solve (session 9). See *Methods* and the Supplemental Materials for more details.

other method. Study methods 1–6 were defined for this experiment as having low effectiveness (based on Dunlosky *et al.*, 2013) and termed here “passive”; study methods 7–17 often include retrieval practice and greater physical activity and were defined for this experiment as having higher effectiveness and termed here “active” (for more discussion on passive and

active study methods see Ward and Walker, 2008; Barger, 2012; Dunlosky *et al.*, 2013; Husmann *et al.*, 2016).

Characteristics of Groups across Experiments

Across the participant groups, demographics were similar but not identical (Table 1). The groups in all three experiments were similar in the perceived amount of studying relative to other students. There were fewer STEM majors among participants in experiment 2 than in other experiments and fewer students of color in the comparison group in experiment 3 than in the intervention group in experiment 3. With these sample sizes and the limited data about participants, we cannot conclude that the groups were equivalent.

Samples and Statistical Analysis

Data analysis was carried out using RStudio 0.97.449 running R 2.14.0 on a Macintosh computer. In all assessments of recall and problem solving, participants were asked to indicate whether they might have made errors in following the methods, and if so, to explain those errors. Data from those known to have made errors following the methods were removed from the data for each quiz before analysis. Sample sizes provided include only those who reported following instructions correctly. However, analyses conducted without excluding these individuals produced qualitatively identical results.

In experiment 1 ($N = 33$), because of nonnormal distributions and unequal variance between groups, data were analyzed using the bootstrap with analysis of variance (ANOVA; Xu *et al.*, 2013a,b). In experiment 2 ($N = 23$), scores were compared using repeated-measures ANOVA;

the factors were language and content page, with study method as the repeated measure. A G -test assessed whether a greater proportion of students scored better with VR or MSFL. In experiment 3 ($N = 36$ –51 in the various quizzes), scores were compared as in experiment 2. Self-scores were compared with investigator scores of recall and problem solving using 1) two-way

TABLE 1. Demographics of participants in surveys and experiments

Experiment	Subgroup	Source	<i>N</i>	Age	Sex (% F)	Students of color (%)	Freshmen (%)	STEM majors (%)
1		Organic Chemistry I	33	18–21	88	36	38	84
2		Introductory Calculus II	23	18–21	61	39	61	57
3	Comparison Pre	Introductory Biology I	55	18–21	75	22	73	75
3	Comparison Post	Introductory Biology I	57	18–21	82	18	59	72
3	Intervention Pre	Introductory Biology I	69	17–21	72	40	74	72
3	Intervention Post	Introductory Biology I	44	18–20	77	40	74	84

repeated-measures ANOVA, with the factors for each study method being content page and language (quizzes 1–3) or biology content (quizzes 4–6), with self-scoring versus investigator-scoring as the repeated measure; and 2) correlation between self-scores and investigator scores using Spearman's correlation coefficient. Survey responses were compared using *t* tests, paired for pre–post in the treatment group (in which the same individuals participated in both surveys) and unpaired for other comparisons. In experiments 2 and 3, when preliminary analysis indicated no significant effect of language or page of content, these factors were not included in the final analyses.

In cases in which data did not meet assumptions of normality or equality of variance, analyses used the bootstrap with *t* tests (Konietzschke and Pauly, 2014) or with ANOVA (Xu *et al.*, 2013a,b). In all cases in which multiple statistical tests were conducted on a single question, the false discovery rate control (Thissen *et al.*, 2002) was used to adjust significance, with the experiment-wise likelihood of acceptance of a falsely significant result set at $p < 0.05$.

Experiments were approved by the College of William and Mary Protection of Human Subjects Committee under protocols PHSC-2014-09-09-9733-pdheid, PHSC-2015-03-23-10285-pdheid, and PHSC-2015-01-27-10051-pdheid.

RESULTS

Experiment 1. Between-Subjects Test for Effectiveness for Recall

The group studying using MSFL had scores for recall that were 20% higher than those for the group studying with VR (Figure 5, $p = 0.046$). The difference was significant despite a

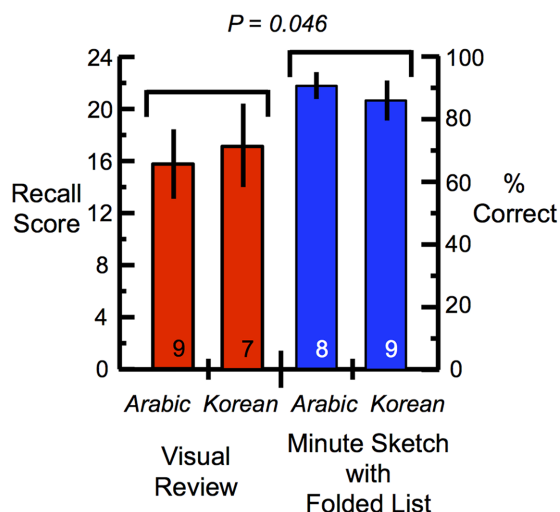


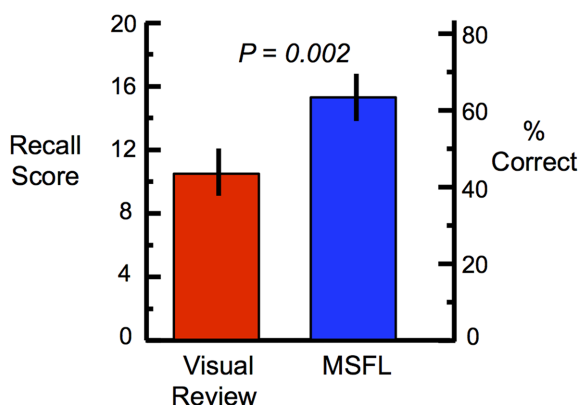
FIGURE 5. Recall score of characters from an unfamiliar alphabet, Arabic or Korean, and associated phonetic spellings of pronunciations by first-year chemistry students after three time-matched study sessions using VR ($N = 16$; red bars) or MSFL ($N = 17$; blue bars). Sample sizes for each alphabet ($N = 7–9$) are indicated within the bars. There was no significant effect of the alphabet studied (lines above bars for each study method); the difference between study methods is indicated by the attained level of significance. A caveat is that the two treatment groups are not known to be equivalent.

ceiling effect, with many perfect or near-perfect scores in the MSFL group. There was no significant effect of language ($p = 0.96$) or interaction between language and study method ($p = 0.61$).

Experiment 2. Within-Subjects Test for Effectiveness for Recall

Unlike the results of experiment 1, in this task with twice the content of experiment 1, only a few individuals achieved high scores of greater than 90% (Figure 6). On average, participants recalled 50% more of the content they studied with MSFL in comparison with VR ($p < 0.01$; Figure 6A). Sixteen of 23 participants (69%) recalled more of the content studied using MSFL

(A) Recall of characters with associated phonetic sounds



(B) Proportion of participants recalling more using a specific study method

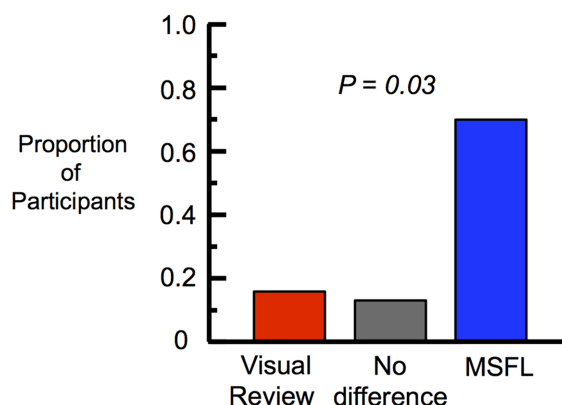


FIGURE 6. Recall by first-year calculus students ($N = 23$) after three time-matched study sessions studying two sets of characters from an unfamiliar alphabet, Arabic or Korean, with associated phonetic spellings of pronunciations. Each participant studied one set of characters using VR (red bar) and a second set of characters using MSFL (blue bar). Differences between alphabets were not statistically significant and are not shown. (A) Scores for recall with attained level of significance. (B) The proportion of participants with higher scores for VR, no difference between methods, or higher scores for MSFL, with attained level of significance for independence.

than the content studied using VR ($p = 0.03$), three were not different (13%), and four participants recalled more content using VR (17%) (Figure 6B).

Experiment 3. Within-Subjects Intervention Test for Effectiveness and Adoption of MSFL

Pre Survey. For both intervention and comparison groups, the most common study method before the experiment was VR (50–60% of study time): rereading notes, presentations, or the textbook. In the pre survey the intervention group reported proportionately less time using passive study methods than the comparison group (passive study methods as 63% vs. 80% of study time, respectively; $p < 0.0001$; Figure 7A). In the pre survey, the comparison group reported less willingness than the treatment group to try different study methods (survey question: “I am reluctant to change my study habits, because they have worked very well for me so far”; $p < 0.05$) and less matching of study methods to study tasks (survey question: “In my studying now, I use different study methods and I match my methods carefully to the material and the skills I need to learn”; $p < 0.05$; see survey in the Supplemental Materials). Thus, the volunteers in experiment 3 were drawn disproportionately from students in the course 1) spending a greater proportion of

their study time using active study methods, 2) more willing to change study habits, and 3) more likely to attempt matching study methods to learning tasks. These *a priori* differences between groups limit the conclusions that can be drawn from comparisons of the two groups.

Quizzes. In self-scored assessment of recall, greater than 75% of participants recalled more with MSFL than with VR in each assessment ($p < 0.01$ for each assessment). On the biology content, there was a weak but statistically significant effect on recall of content set A versus B ($p < 0.05$). In self-scored assessments based on the ratio of differences, participants recalled an average of 68% more of the content they studied with MSFL in comparison with VR ($p < 0.01$; Figures 8A and 9A). Self-scores and investigator scores were highly correlated (characters, $R = 0.912$, $p < 0.0001$; biology content, $R = 0.82$, $p < 0.0001$). On the basis of scoring by investigators, participants recalled an average of 48% more when using MSFL in comparison with VR (Figures 8B and 9B). In individual outcomes, 81% of participants recalled more with MSFL than with VR, 5% had identical recall using the two methods, and 14% recalled more using VR (Figure 10). Results were quantitatively similar and qualitatively identical

in analyses broken down by sex, age (grouped as ages 17–18 and 19–20), ethnicity (grouped into underrepresented minorities, Asian, or white), self-assessed study time in relation to other students (grouped into those who felt they study more than other students or not more than other students), and willingness to change study methods (grouped as reluctant to change or willing to change study methods). In all these groupings, recall using MSFL was greater than when using VR, with all p values 0.01 or smaller, with no significant effect across subgroups.

Self-scores for problem solving were poorly correlated with investigator scores, so only investigator scores were used to assess problem solving. Scores for solving problems on the biology content studied using MSFL were significantly better than VR on quiz 4 (study method: $F = 8.33$; $p = 0.007$; content page: $F = 1.11$; $p = 0.30$; interaction: $F = 1.07$; $p = 0.31$), while on quiz 5, after 2 weeks with no studying, there were no significant factors (study method: $F = 1.25$; $p = 0.27$; content page: $F = 3.39$; $p = 0.08$; interaction: $F = 0.70$; $p = 0.41$). On quiz 6, after a single additional study session, there was a significant interaction term, indicating better problem solving using MSFL in one of two pages of content (interaction between study method and content page: $F = 4.36$; $p = 0.04$; study method: $F = 1.12$; $p = 0.30$; content page: $F = 1.51$; $p = 0.23$). In the final problem-solving session, session 9, there was no significant effect of MSFL on results

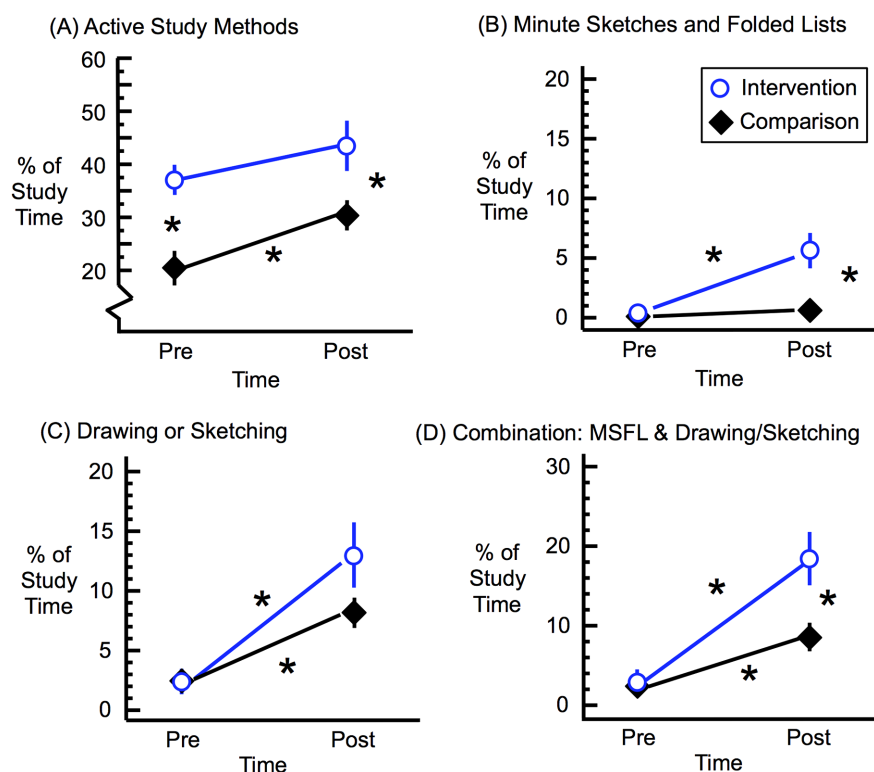


FIGURE 7. Self-reported study time devoted to different study methods at the beginning of the study (September 2014) and the end of the academic year (April 2015) for all individuals in the intervention group (open circles, blue) and those in a comparison survey drawn from the same biology course (closed diamonds, black) showing the percentage of study time (A) using active methods, (B) including MSFL, (C) using other methods incorporating drawing or sketching, and (D) including the combined use of MSFL and drawing or sketching. In each figure, significant differences between groups are indicated by asterisks ($p < 0.05$; adjusted by the false discovery rate control; see *Methods*).

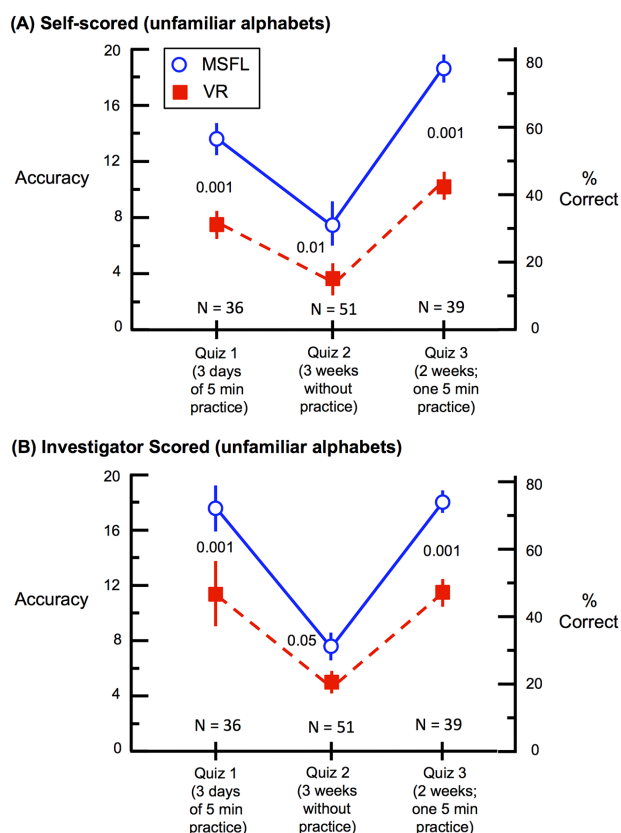


FIGURE 8. Recall by first-year biology students after time-matched study sessions studying two sets of characters from an unfamiliar alphabet, Arabic, Korean, or Tamil, with associated phonetic spellings of pronunciations. Each participant studied one set of characters using VR (solid squares, red) and a second set of characters from the same, unfamiliar alphabet using MSFL (open circles, blue), randomly assigned to a specific study method for character set A. Participants were quizzed three times (see Figure 4 for timeline). Quiz 1 followed 3 days with practice, quiz 2 occurred after an additional 3 weeks with no practice, and quiz 3 was given after an additional 2 weeks with a single practice session the day before the quiz. The attained level of statistical significance is indicated for each quiz. Sample sizes for each quiz are indicated above the x-axis. (A) Self-scores by participants, each scoring his or her own quiz using a simple rubric supplied by the investigators. (B) Scoring conducted by investigators, blind with respect to study method and participant ID. Because there were no statistically significant effects of the language of the alphabet or content page, these two factors are collapsed in this figure.

of the problem-solving task ($p > 0.10$ for all tests). Overall, MSFL resulted in improved problem solving in time-matched testing in comparison with the preferred study method (VR) in some assessments, but not the majority. In no assessment did VR provide better problem solving than MSFL.

For study using VR, our participants reported two methods: 1) systematically looking at and trying to impress content into memory and 2) repeatedly looking away or closing eyes to practice recall, followed by looking back to check accuracy (i.e., VR with retrieval practice). These two methods of VR did not differ in effectiveness: participants who reported method 1 or 2 did not differ in the amount recalled ($p > 0.50$).

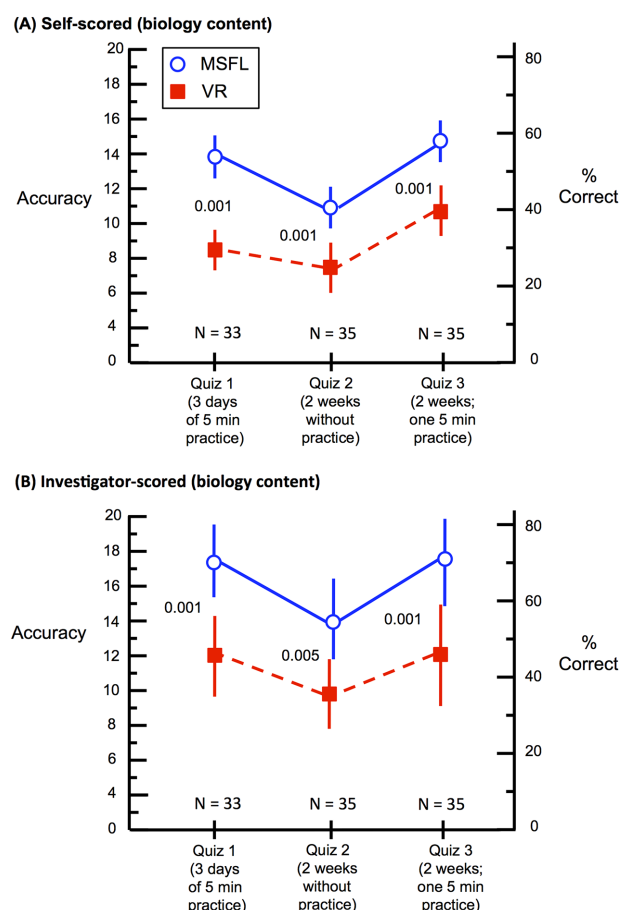


FIGURE 9. Recall by first-year biology students after time-matched study sessions studying two sets of unfamiliar biology content (thyroid hormone structure and function and regulation of metabolism). Each participant studied one set of content using VR (solid squares, red) and the second set of content using MSFL (open circles, blue), randomly assigned to a specific study method for content set A. Participants were quizzed three times (see Figure 4 for timeline): after a week including 3 days with practice (quiz 1), after an additional 2 weeks with no practice (quiz 2), and after an additional 2 weeks with a single practice session (quiz 3). The attained level of statistical significance is indicated for each quiz. Sample sizes for each quiz are indicated above the x-axis. (A) Self-scores by participants, each scoring his or her own quiz using a rubric supplied by the investigators. (B) Scoring by investigators, blind with respect to study method and participant ID.

In the post survey, many participants in the intervention group reported their continued use of minute sketches and/or folded lists (Figure 11). In the pre-intervention survey, reported use of minute sketches and folded lists was ~2% of study time, and did not differ between the intervention group and comparison group (Figure 7B; Fisher exact test; $p > 0.1$ for both). In the post survey 6 months after introduction to the method, approximately two-thirds of participants in the intervention reported use of minute sketches or folded lists (Figure 11), with a significant increase in comparison with the pre survey ($p < 0.05$; Figure 7B). In addition, the intervention group was significantly more likely to use MSFL

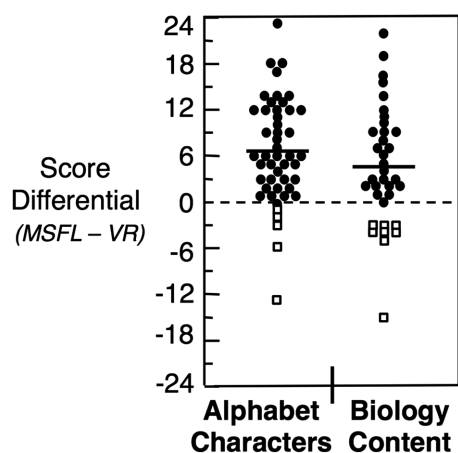


FIGURE 10. Score differentials for study method (data from experiment 3), showing the distribution of individual outcomes. Solid circles represent participants with equal or greater recall using MSFL, and open squares indicate those with greater recall using VR.

or other drawing and sketching than the comparison group ($p < 0.05$; Figure 7D).

We explored whether continued participation through half or more of the intervention was related to increases in the use of drawing or sketching during studying, to the adoption of MSFL, or to changes in the use of active study methods. In the pre survey, students in the intervention and comparison groups made minimal use of drawing or sketching (average: 2% of study time; Figure 7). Individuals participating only in the first three or four sessions of the intervention did not change significantly between the pre and post surveys (Figure 12). In contrast, individuals who participated in more than half of the intervention, at least through the first quiz using biology content, differed significantly in 1) the percent of time allocated to active study methods, 2) use of MSFL, 3) drawing or sketching apart from MSFL, and 4) the combination of MSFL with other drawing and sketching (Figure 12). Individuals participating for more than half of the intervention were those most likely to report continued use of MSFL (averaging 8% of study time; Figure 12B) and increased their use of all forms of drawing and sketching (averaging 24% of study time) in comparison with the individuals who dropped from the intervention after the early sessions (Figure 12, C and D) and the comparison group (compare Figures 7 and 12).

DISCUSSION

This study assessed MSFL as a study method in comparison with a preferred study method, VR, and taught the method to students in an intervention and allowed them to self-assess the effectiveness of the method. Any learning gain that motivates students to use drawing as a learning strategy will create opportunities for development of higher-order skills with drawing, supporting an ultimate goal of students generating drawings that they use for purposes ranging from recall to complex problem solving. To overcome student reluctance to adopt a new study method (Pressley *et al.*, 1989; NRC, 2001; Fu and Gray, 2004; Dunlosky *et al.*, 2013), we focused the attention of

participants on two potential benefits, better recall and better problem solving. Our results suggest that many participants perceived gains in learning from MSFL and continued to use drawing and/or MSFL as a study strategy in subsequent months.

Participants using MSFL retained ~50% more content in time-matched comparisons with a preferred study method, regardless of whether the content was shapes and sounds (experiment 1, Figure 5; experiment 2, Figure 6A; experiment 3, Figure 8) or new content in biology (experiment 3, Figure 9). The effect was maintained in multiple tests over a period of 4–5 weeks (experiment 3, Figures 8 and 9). Importantly, student self-scores for recall were highly correlated with instructor scores (see *Results* and Figures 8 and 9), suggesting that student self-scores were reasonably reliable assessments of gains. Approximately 80% of participants recalled more with MSFL than when using VR (Figure 10). Even though instruction in problem solving using minute sketches was a minor part of the intervention, there was some evidence for improved problem solving after study using MSFL. Overall, when study time was held constant, MSFL enhanced problem solving in some trials and enhanced recall in all trials. Participants in the intervention were likely to adopt elements of MSFL, particularly drawing and sketching, in later studying (Figures 7, 11, and 12).

Because an observer effect may cause participants to meet perceived goals of investigators, we conducted one experiment in which each participant was presented with only one of the two methods (experiment 1, Figure 5). Participants in experiment 1 were given no information on the potential effectiveness of study methods and might reasonably have inferred that the experimental activity compared the difficulty of learning two different alphabets. There was an unanticipated ceiling effect, with two-thirds in the MSFL treatment achieving perfect or near-perfect recall of the single set of 12 characters and 12 associated sounds. Nonetheless, studying using MSFL outperformed VR (Figure 5). In experiment 2, in which each participant applied both study methods but received no information on effectiveness of different study methods, recall after study using MSFL again outperformed VR (Figure 6). While there were potential differences in group composition in experiment 1 and small sample sizes in experiments 1 and 2 (see *Limitations* section), these results suggest that greater recall when using MSFL may be independent of information or expectations given to participants regarding effectiveness. The results are consistent with multiple studies showing benefits to recall from associating simple drawings with words (see Wammes *et al.*, 2016).

Our second objective was to test whether an intervention (experiment 3) using MSFL with self-assessment of effectiveness might be followed by adoption of MSFL in later studying. The intervention first introduced students to the study method, allowed them to self-assess gains on a simple memory task, and then allowed them to self-assess gains on recall and problem solving with biology content. Six months after the beginning of the intervention, a majority of survey respondents reported adopting MSFL (Figure 11). Of those completing more than half of the intervention in experiment 3, involving 4.5–8.5 hours of participant time, the proportion of study time using some form of drawing or sketching increased significantly in the intervention group (Figure 12, C and D). In contrast, participants who ended participation before the assessments on

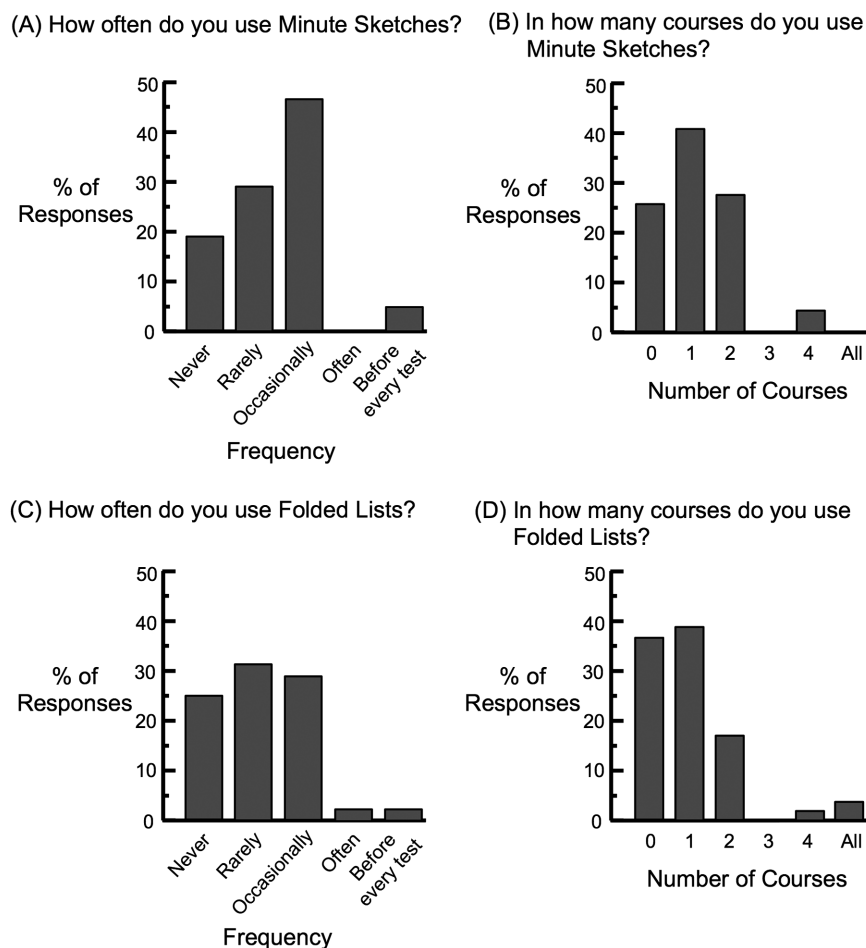


FIGURE 11. Reported use of MSFL by individuals in the intervention group when surveyed 6 months after initial instruction. ($N = 44$ of the 56 participants who completed at least two sessions for the intervention.)

biology content beginning in session 5 were 1) unlikely to adopt MSFL (Figure 12B) and 2) fairly similar to the comparison group in their post survey responses on MSFL, drawing, and sketching (compare Figure 7, B–D, with Figure 12, B–D). Continued participation through at least session 5 may have increased the likelihood of adopting MSFL. It is also possible that an early decision to adopt MSFL increased the likelihood of continuing participation in the study. Regardless of the cause, a high proportion of participants had adopted MSFL in their studying, with combined use of drawing and sketching reported at nearly 25% of study time for those participating in more than half of the intervention.

Minute sketches are intended to be useful for problem solving, a goal supported by studies on drawing-to-learn (Van Meter and Garner, 2005; Quillin and Thomas, 2015). After first learning the biology content, students were better at problem solving on content they had learned using MSFL. After 2 weeks without studying, the difference in problem solving was absent (Figure 9), perhaps because content learned with both methods was not remembered well. Two weeks later, after a single review session, problem solving was better for one but not both pages of the content studied using MSFL. Thus, studying content using MSFL

gave significantly better problem-solving outcomes 1) one to several days after learning and 2) after review. This might be evidence that minute sketches were being used for problem solving, but a likely alternative is that differences may have been based on the differences in recall: it is difficult to solve problems on content that is not recalled. In the final part of the intervention, in which a small subsample of participants were given instructions on the use of minute sketches as visual models to solve problems, there was no difference in problem solving on content studied using MSFL or using VR. If minute sketches are a format for learner-generated visual models that may be useful for problem solving, it may be that most students need more instruction on problem solving than provided in this study (see Van Meter and Garner, 2005; Schwarz *et al.*, 2009; Quillin and Thomas, 2015). More research is needed to address this question.

Greater recall using MSFL in comparison with VR is unsurprising, because MSFL involves practice testing and retrieval practice, approaches to studying that are more effective than VR with rereading (Karpicke and Blunt, 2011; Dunlosky *et al.*, 2013). Drawing or sketching alone can improve recall and understanding and help in solving problems (Van Meter and Garner, 2005; Ainsworth *et al.*, 2011; Quillin and Thomas, 2015; Wammes *et al.*, 2016). Nonetheless, useful study methods misaligned to a particular learning task might decrease learning (Leutner *et al.*, 2009; Schwaborn *et al.*, 2011; Leopold and Leutner, 2012). On the basis of our results (Figures 4, 6, 8, 9, and 10), MSFL may be a useful study tool for college students in biology as a replacement for a common and preferred study method, VR. It may be important that MSFL be presented as in our intervention, in which students learn the method and self-assess MSFL in comparison with a common preferred study method such as VR. In this study, presentation as an intervention requiring at least 4.5 hours and up to 8.5 hours of participant time was associated with a substantial increase, 6 months later, in the use of MSFL and other forms of drawing/sketching, to ~25% of study time. The results suggest that this intervention may overcome resistance of students to changes in study behavior (Pressley *et al.*, 1989; NRC, 2001; Fu and Gray, 2004). Nonetheless, application of drawing for more complex tasks may require additional instruction and practice by students (Van Meter and Garner, 2005; Schwarz *et al.*, 2009; Quillin and Thomas, 2015).

LIMITATIONS

While results were consistent across experiments, sample sizes were not large, and we did not assess participant features such as prior academic ability, course achievement, and proficiency

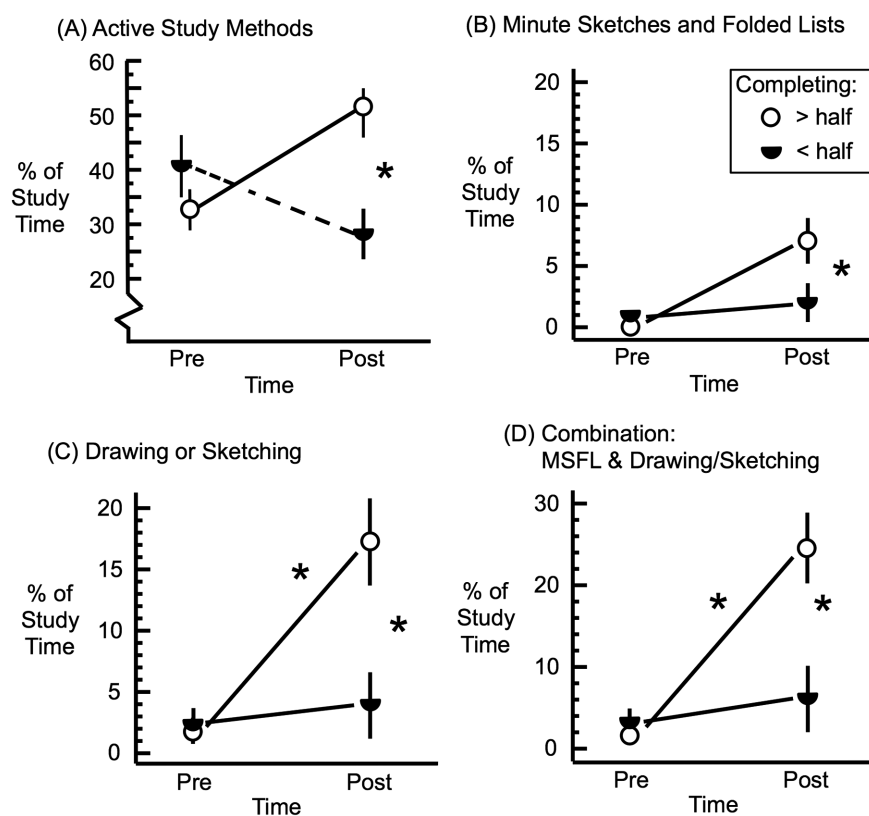


FIGURE 12. Comparison of self-reported study time of two subsets of participants within the intervention group in the pre survey (9/2014) and post survey (4/2015). The two subgroups comprised those completing at least two quizzes, but less than half of the study (three to four sessions; $N = 10$; closed half-circles) and those completing at least half of the study (five or more sessions; $N = 23$; open circles). The panels show the percentage of study time (A) using active methods, (B) using MSFL, (C) using other methods incorporating drawing or sketching, and (D) for the combined use of MSFL and drawing or sketching. In each figure, asterisks (*) indicate differences significant at $p < 0.05$ following assessment using the false discovery rate control (see *Methods*).

with languages or biology. Therefore, there may be other contributing variables that explain differences in performance on our assessments. In addition, the surveys and quizzes have not been validated to assess how well they gather the information intended. Nevertheless, our study is a starting point for future research on learning and problem solving after practice with quickly reproduced “minute sketches” with associated terms that capture, for a user, a structure, concept, or series of events.

CONCLUSION

We view this intervention as a practical method to introduce students to drawing as a study tool, ideally taught along with information on metacognition and self-regulated studying. MSFL is designed to have simple instructions that students learn as part of the intervention, meeting a need for instruction and practice for effective application of drawing to learning (Quillin and Thomas, 2015). The initial learning task was simple—characters with sounds—and analogous to learning tasks students often face in STEM: learning new symbols (such as μ or Å) or letters that serve as symbols (such as Na for sodium or Hg for mercury) and pronunciations, a routine task

in many content areas in STEM. In fact, if this intervention was part of a course, the initial learning task could be used to begin teaching new symbols. The discovery by students that MSFL can lead to faster mastery of learning tasks (sessions 2–4), however simple, could motivate students to continue applying the study method. The second learning task in the intervention, unfamiliar discipline-based content, allows students who have practiced the method a second chance to assess their learning, this time as applied to discipline-based content (sessions 5–7). Again, the discovery that MSFL can lead to better recall could motivate students to apply MSFL, or drawing more generally, to additional content. While students showed some gains in problem solving, it seems likely that, consistent with other reports, skill at applications of drawing for higher-order learning might require continued practice with instruction (Van Meter and Garner, 2005; Schwarz *et al.*, 2009; Quillin and Thomas, 2015). Overall, these results suggest that this intervention, or modifications of this intervention adapted to specific courses, could lead to adoption of drawing by students at least for lower-order learning tasks that might set the stage for broader application of learner-generated visual models. Even without further practice and instruction, students applying MSFL may show improvements in problem solving. If learner-generated drawing is a core skill in biology (Quillin and Thomas, 2015), interventions that motivate students to apply drawing as a long-term

study strategy will be useful to develop that skill.

ACKNOWLEDGMENTS

Funding was provided by National Science Foundation DUE 1339939 (a Robert Noyce Teacher Scholarship Program grant to the College of William and Mary) and a Howard Hughes Medical Institute Undergraduate Science Education grant to the College of William and Mary. Invaluable assistance in experimental design was provided by the Biology Scholars Program of the American Society for Microbiology, particularly by M. A. Kelly, C. P. Davis, L. Clement, A. Hunter, P. Soneral, K. Wester, and M. Zwick. P. Soneral provided insightful comments and suggestions on an earlier draft of the manuscript. The William and Mary Noyce Scholars in Fall 2012 and 2013 suggested the use of Korean and Arabic characters and helped develop the experimental protocol for assessment when learning characters. For particularly helpful discussion and feedback in the development of the learning method, we thank S. L. Sanderson, M. S. Saha, R. H. Macdonald, J. J. Matkins, M. M. Mason, K. D. Goff, and C. Walck. We also thank two anonymous reviewers for extremely helpful comments and suggestions.

REFERENCES

- Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, 333, 1096–1097.
- Barger, J. B. (2012). How do undergraduate students study for anatomy, and does it matter? *FASEB Journal*, 26, 528–522.
- Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13, 221–252.
- Bransford, J. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.
- Chang, C. Y., Yeh, T. K., & Barufaldi, J. P. (2010). The positive and negative effects of science concept tests on student conceptual understanding. *International Journal of Science Education*, 32, 265–282.
- Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50, 639–659.
- Dowd, J. E., Duncan, T., & Reynolds, J. A. (2015). Concept maps for improved science reasoning and writing: complexity isn't everything. *CBE—Life Sciences Education*, 14, ar39.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58.
- Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: sketch understanding for cognitive science research and for education. *Topics in Cognitive Science*, 3, 648–666.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, 111, 8410–8415.
- Fu, W.-T., & Gray, W. D. (2004). Resolving the paradox of the active user: stable suboptimal performance in interactive tasks. *Cognitive Science*, 28, 901–935.
- Hay, D. B., Williams, D., Stahl, D., & Wingate, R. J. (2013). Using drawings of the brain cell to exhibit expertise in neuroscience: exploring the boundaries of experimental culture. *Science Education*, 97, 468–491.
- Husmann, P. R., Barger, J. B., & Schutte, A. F. (2016). Study skills in anatomy and physiology: Is there a difference? *Anatomical Sciences Education*, 9, 18–27.
- Jee, B. D., Gentner, D., Uttal, D. H., Sageman, B., Forbus, K., Manduca, C. A., ... Tikoff, B. (2014). Drawing on experience: how domain knowledge is reflected in sketches of scientific structures and processes. *Research in Science Education*, 44, 859–883.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772–775.
- Konietschke, F., & Pauly, M. (2014). Bootstrapping and permuting paired t-test type statistics. *Statistics and Computing*, 24, 283–296.
- Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, 22, 16–26.
- Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior*, 25, 284–289.
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge, UK: Cambridge University Press.
- McDermott, K. B. (2006). Paradoxical effects of testing: repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, 34, 261–267.
- National Research Council (NRC). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.
- NRC. (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.
- Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, 13, 197.
- Pressley, M., Goodchild, F., Fleet, J., Zajchowski, R., & Evans, E. D. (1989). The challenges of classroom strategy instruction. *Elementary School Journal*, 89, 301–342.
- Quillin, K., & Thomas, S. (2015). Drawing-to-learn: A framework for using drawings to promote model-based reasoning in biology. *CBE—Life Sciences Education*, 14, es2.
- Roediger, H. L. III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1155–1159.
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19, 469–508.
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36, 111–139.
- Schwaborn, A., Thillmann, H., Opfermann, M., & Leutner, D. (2011). Cognitive load and instructionally supported learning with provided and learner-generated visualizations. *Computers in Human Behavior*, 27, 89–93.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46, 632–654.
- Speth, E. B., Shaw, N., Momsen, J., Reinagel, A., Le, P., Taqieddin, R., & Long, T. (2014). Introductory biology students' conceptual models and explanations of the origin of variation. *CBE—Life Sciences Education*, 13, 529–539.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185–233.
- Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, 11, 113–120.
- Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27, 77–83.
- Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review*, 17, 285–325.
- Verhoeven, L., Schnotz, W., & Paas, F. (2009). Cognitive load in interactive knowledge construction. *Learning and Instruction*, 19, 369–375.
- Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, 69, 1752–1776.
- Ward, P. J., & Walker, J. J. (2008). The influence of study methods and knowledge processing on academic success and long-term recall of anatomy learning by first-year veterinary students. *Anatomical Sciences Education*, 1, 68–74.
- Wright, L. K., Fisk, J. N., & Newman, D. L. (2014). DNA→RNA: What do students think the arrow means? *CBE—Life Sciences Education*, 13, 338–348.
- Xu, L.-W., Mei, B., Chen, R.-R., Guo, H.-X., & Wang, J.-J. (2013a). Parametric bootstrap tests for unbalanced nested designs under heteroscedasticity. *Journal of Statistical Computation and Simulation*, 84, 2059–2070.
- Xu, L.-W., Yang, F.-Q., Abula, A., & Qin, S. (2013b). A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances. *Journal of Multivariate Analysis*, 115, 172–180.