

RESEARCH ARTICLE

A dissociation between engagement and learning: Enthusiastic instructions fail to reliably improve performance on a memory task

Benjamin A. Motz^{1,2*}, Joshua R. de Leeuw³, Paulo F. Carvalho⁴, Kaley L. Liang¹, Robert L. Goldstone^{1,2}

1 Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, United States of America, **2** Cognitive Science Program, Indiana University, Bloomington, Indiana, United States of America, **3** Cognitive Science Department, Vassar College, Poughkeepsie, New York, United States of America, **4** Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America

* bmotz@indiana.edu



OPEN ACCESS

Citation: Motz BA, de Leeuw JR, Carvalho PF, Liang KL, Goldstone RL (2017) A dissociation between engagement and learning: Enthusiastic instructions fail to reliably improve performance on a memory task. *PLoS ONE* 12(7): e0181775. <https://doi.org/10.1371/journal.pone.0181775>

Editor: Etsuro Ito, Waseda University, JAPAN

Received: May 25, 2017

Accepted: July 6, 2017

Published: July 21, 2017

Copyright: © 2017 Motz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All materials, including videos, raw data, and analysis scripts, are available at <http://osf.io/j2h8y/>.

Funding: This research was supported in part by supplemental funding from the Indiana University Department of Psychological and Brain Sciences, a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1342962 to JRD, and a Graduate Training Fellowship SFRH/BD/78083/2011 from the Portuguese Foundation for Science and Technology (FCT), co-sponsored by the European Social Fund, to PFC. The funders had

Abstract

Despite widespread assertions that enthusiasm is an important quality of effective teaching, empirical research on the effect of enthusiasm on learning and memory is mixed and largely inconclusive. To help resolve these inconsistencies, we conducted a carefully-controlled laboratory experiment, investigating whether enthusiastic instructions for a memory task would improve recall accuracy. Scripted videos, either enthusiastic or neutral, were used to manipulate the delivery of task instructions. We also manipulated the sequence of learning items, replicating the spacing effect, a known cognitive technique for memory improvement. Although spaced study reliably improved test performance, we found no reliable effect of enthusiasm on memory performance across two experiments. We did, however, find that enthusiastic instructions caused participants to respond to more item prompts, leaving fewer test questions blank, an outcome typically associated with increased task motivation. We find no support for the popular claim that enthusiastic instruction will improve learning, although it may still improve engagement. This dissociation between motivation and learning is discussed, as well as its implications for education and future research on student learning.

Introduction

Enthusiasm is widely believed to be fundamental to effective teaching. Expert teachers routinely promote the crucial importance of enthusiastic instruction for inspiring student interest, motivation, and attention toward the course topics and activities (e.g., [1–3]). In his early survey of teaching behaviors, Barr [4] observed that enthusiastic “energy and vitality” were the principal differentiating characteristics of “good” teachers. In McKeachie’s popular guidebook of teaching tips, the lecturer’s primary role in higher education is reduced simply: “to communicate the teacher’s enthusiasm for the subject.” ([5], p. 71). Undergraduate students also rank enthusiasm among the most important and valuable properties of instructors [6, 7], and accordingly, student ratings of teachers’ enthusiasm represent a dominant factor in end-of-

no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

semester evaluations [8, 9]. Moreover, instructors who receive uniformly high marks on student evaluations also tend to rank particularly high on extraversion personality scales, marked by behavioral traits such as being excited, outgoing, and playful [10, 11].

One might assume that the importance of enthusiasm, as assessed by expert teachers and students alike, would be supported by strong evidence of its relationship to student learning. However, in contrast to these strong intuitions, research results are mixed. Even while enthusiastic teaching reliably correlates with student interest, involvement, and enjoyment (reviewed in [12]; c.f. [13]), a clear connection between enthusiasm and student achievement (test performance, course grades, etc.) has not been identified.

Mastin [14] conducted the earliest cited experimental manipulation of teacher enthusiasm. A teacher gave two lecture-style lessons to 6th and 7th graders, one-week apart, regarding two different ancient civilizations. One lesson was delivered in an indifferent manner, and the other in an enthusiastic manner, counterbalanced for order of presentation and the civilization described. On a subsequent memory test, 19 of the 20 classrooms had better performance scores for the civilization that had been presented enthusiastically. Additional published experiments have similarly observed improved learning outcomes under enthusiastic teaching conditions [15–19], however, some published studies have also failed to observe these effects [20–23], and some suggest that any effect of enthusiasm may become negligible when students are given an extrinsic incentive to learn the material [24, 25]. Interpretation of these mixed results is difficult because many of these studies lacked random assignment of learners to classroom experimental conditions [15, 17, 20–24], or because additional factors (such as the amount of instructor-student interaction) were comingled in the enthusiasm condition [19, 26]. Unfortunately, experimental attempts that included careful random assignment of students to controlled enthusiasm conditions have also observed no significant effects, or very limited effects, of enthusiasm on learning [27–29].

These mixed results have been observed under a range of operationalizations of enthusiasm, from a specific set of nonverbal gestures [30], to broader conceptualizations including the expression of humor, energy, and charisma [31], such that inconsistent findings cannot be conveniently attributed to semantic differences in how different researchers define or manipulate enthusiasm. Effects on student achievement, or lack thereof, are not limited to some narrow set of instructor actions or behaviors. This should make sense because different expressions of enthusiasm are not independent from each other. Indeed, factor analyses have observed that expressive qualities tend to covary in authentic instruction [9, 32], and for this reason, researchers have suggested that inclusive and multifaceted definitions of enthusiasm are preferable [12, 33].

Theoretically, the effect of enthusiastic instruction should be to improve a student's motivation to do something [34]. But past experimental research on enthusiasm has emphasized the teacher's enthusiastic delivery of *information* to be learned, and not the enthusiastic delivery of *instructions* for a learning activity. Considering that self-regulated learning is playing an increasingly vital role in educational systems, enthusiastic instruction should motivate a student's independent studying behaviors, not merely animate the to-be-learned material. That is, for situations in which students will be assigned to seek out their own individualized learning materials, instructors will often be in the position of being able to vary their enthusiasm in exhorting the student to go forth and learn, but will not be able to directly affect the enthusiasm exhibited in the materials themselves. In support of this perspective, there have been some indications that a teacher's enthusiasm will only affect memory performance when students consider themselves to be in control of their own learning [35, 36].

The rather inconclusive state of research on instructor enthusiasm stands in stark contrast to recent developments in cognitive learning interventions. Memory researchers have amassed

a considerable catalogue of studying techniques that improve learning, and generalize to a range of educational applications [37]. For example, the *spacing effect* (memory improvement when repetition of study items is distributed over time rather than massed; [38]), one of the largest and most reliable effects in the learning and memory literature [39], has been repeatedly demonstrated to improve learning and student achievement in authentic educational contexts (e.g., [40–45]). But these cognitive techniques, despite their reliability and effectiveness, remain relatively unknown to educators [46], and students often do not recognize their practical value [47, 48]. Enthusiasm, on the other hand, may be a widespread pedagogical stratagem in education, but its relationship to student learning lacks for unequivocal empirical support.

The lack of consistent findings may be, in part, because past researchers have aimed to quantify enthusiasm's effect in naturalistic educational settings. While these efforts toward ecological validity are laudable [49, 50], the current study instead aims to quantify the effect of instructor enthusiasm in a delayed-recall laboratory memory experiment. Advantages of this approach include the ability to randomly-assign individual participants to different instructional conditions (thus avoiding sampling bias of classroom groups, as well as the potential for learner-to-learner interactions that are not of interest to the current study), to tightly control the stimuli and learning materials (thus avoiding confounds from teacher-student interactions that might be concomitant with face-to-face enthusiastic teaching), and to emulate the individualized and self-paced instructions that are characteristic of contemporary blended learning environments. And in doing so, the effect of enthusiasm can also be measured alongside the effect of a cognitive manipulation, such as the spacing effect.

Considering that enthusiasm occupies a high pedestal among teaching behaviors, and that the research foundation for its status is not particularly sturdy, the literature on its relationship to student achievement can clearly benefit from a controlled memory experiment. The primary goal of the current study is to address this need, examining whether enthusiastically-delivered instructions can affect performance in a laboratory memory task. Because enthusiasm has been claimed to affect both learning and motivation [51], we focus on two distinct measures of task performance: recall accuracy (the number of study items correctly remembered after a delay) and recall quantity (the total number of responses, correct and incorrect, provided at test; which has been observed to be sensitive to experimental manipulations of motivation and encouragement [52, 53]). We also seek to quantify these effects of instructor enthusiasm alongside the effect of temporally spacing the repetition of study items (the spacing effect), a known cognitive technique for memory improvement. By manipulating instructional enthusiasm in two different sequence conditions, we add greater contextualization to our assessment of the main effect of enthusiasm, given different study tactics. Moreover, this direct contrast between instructional behaviors and study sequencing may help provide a better understanding of the relative efficacy of these two educational strategies for improving learning and memory.

Experiment 1 (Exploratory study)

Experiment 1 was intended to provide a preliminary examination of whether enthusiastically-delivered instructions would affect performance on a typical paired-associates delayed-recall memory task. In order to maximize experimental control of the instructor's delivery, we made two video recordings of the task instructions presented by the same instructor, one was enthusiastic, the other neutral. Participants watched either video just prior to studying the paired associates, one pair at a time, in a computer-guided sequence that either distributed the study items uniformly (spaced) or repeated the study items back-to-back (massed). One week later, participants returned to the lab for a recall test. All materials, including videos, raw data, and analysis scripts, are available at <http://osf.io/j2h8y/>.

Method

Ethics statement. The experimental procedures, materials, and recruitment protocols were approved by the Indiana University Institutional Review Board. All participants provided informed consent electronically at the start of the experiment (see Procedure section below).

Participants. Eighty-six undergraduate students at Indiana University Bloomington volunteered to participate in this two-session study toward completion of their Introductory Psychology experiment participation requirements. Sample size for this exploratory study was not predetermined, and data collection was stopped at the end of the academic term. Participants were informed in the experiment instructions that, in addition to the credits toward their course requirement, they would also be paid a \$0.25 bonus for each word pair correctly recalled (the maximum bonus was \$9) at the end of the second session. Seven participants were excluded from analyses because they did not return to complete the second session ($N = 5$), there was an error during data collection ($N = 1$), or because the participant failed to complete the study phase in the allotted time ($N = 1$); leaving 79 participants in total, 48 females (by self-report; 61%), between the ages of 18 and 24 years ($M = 19.4$, $SD = 1.2$). Sixty-one of these participants reported being native English speakers (77%), and the remaining non-native English speaking participants had been speaking English for an average of 12.3 years ($SD = 3.4$). No Indonesian speakers volunteered to participate in this study.

Stimuli. We selected 36 English-Indonesian word pairs from a larger set of 96 word pairs used in [54]. Based on the norming data provided, the selected words had recall accuracy ranging from 43% to 100%.

Video recorded instructions. We wrote comprehensive task instructions for this memory experiment in two different scripts; one was written so as to demonstrate enthusiasm for the participant's task performance, and the other was neutral. Each of these scripts contained the same structure and basic information, with an instructor initially welcoming the participant to the study, providing background information, describing step-by-step procedures, and so on. The enthusiastic script was written to express charisma, personality, and humor (Williams and Ware, 1976), and accordingly, it included language that was more active, personally relevant, encouraging, and specific. The neutral script contained all the same task directions, but it was written in a more general and impersonal tone than the enthusiastic script, including passive phrases and unadorned straightforward statements and instructions.

We posted a job advertisement to graduate students enrolled in Indiana University's Masters of Fine Arts (MFA) acting program, to recruit a paid performer for these scripts. We ultimately hired a talented female actor, Ashley, who had experience making professional training videos, and also had extensive experience teaching undergraduates. She was asked to deliver the enthusiastic script with varied vocal inflexion, demonstrative gesturing and movements, and emotive facial expressions [30], imagining an enthusiastic and dynamic teacher who is well-liked by her students. She prepared the neutral script without these expressive qualities, imagining a teacher who is polite and speaks clearly, but who isn't expressive or particularly invested in inspiring her students. We described the experiment to the actor, and she confirmed that she would be able to comfortably assume these instructional roles in a realistic manner. She memorized the scripts, and we staged and recorded the instructions in the same testing room that participants would be performing the experiment.

The total duration of all recorded video segments in the enthusiastic condition was 4 minutes 37 sec, and 4 minutes 6 seconds in the neutral condition (see Fig 1 for screenshots; the individuals shown in this figure have given written consent to publish these images).

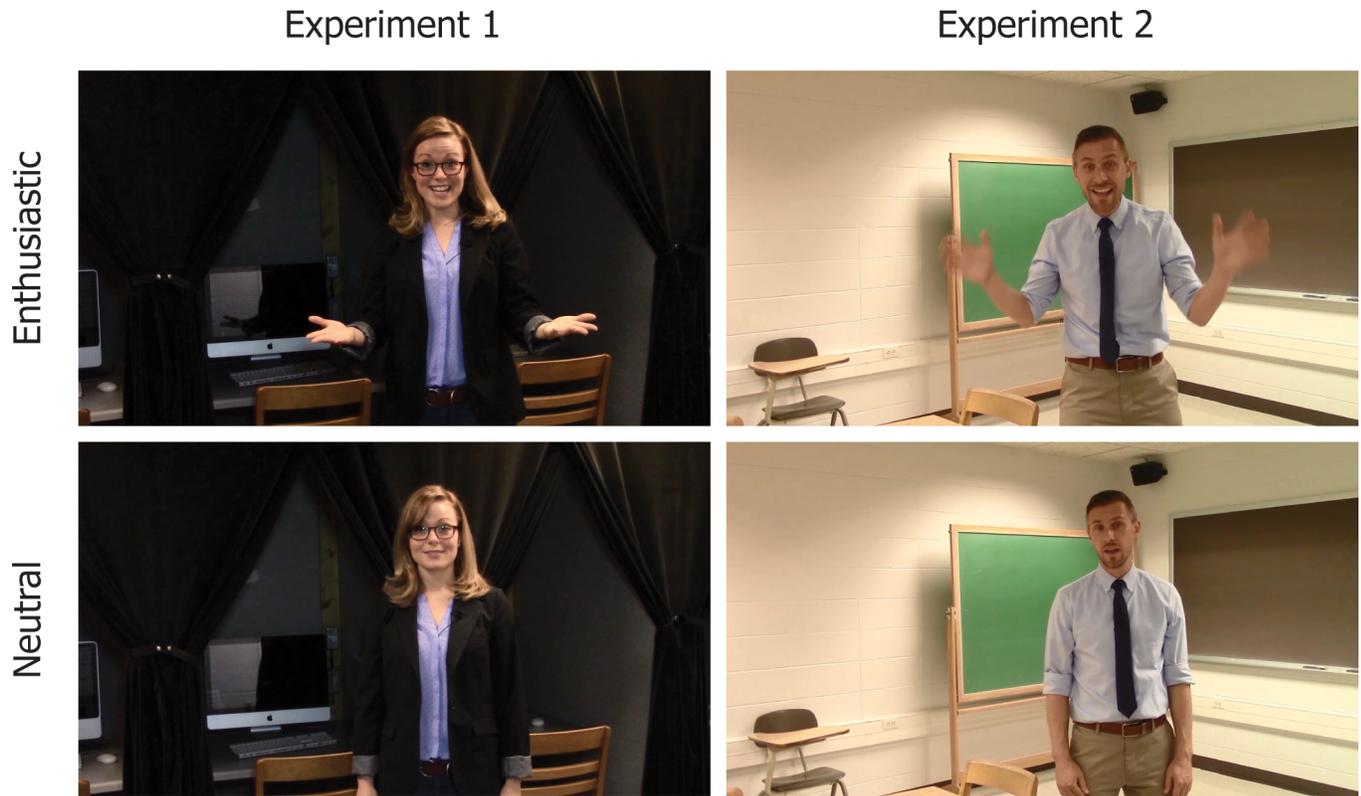


Fig 1. Examples from instructional videos. Representative stills at equivalent moments in enthusiastic and neutral instructional videos. Experiment 1 stills are from the video shown halfway through the study session. Experiment 2 stills are when providing background information about Indonesian at the start of the study session. Full videos available at <http://osf.io/j2h8y/>.

<https://doi.org/10.1371/journal.pone.0181775.g001>

Procedure. When initially signing-up for the experiment, participants scheduled two separate appointments (first for a study session, and then for a recall test session), with the constraint that the second session take place 5 to 9 days after the first session.

When they arrived for the first session, participants were greeted by an undergraduate research assistant, who led them to individual testing cubicles, and initialized the experiment on a desktop computer with comfortable headphones. Upon initializing, the experiment script (written with jsPsych [55]) automatically assigned each participant to an enthusiasm condition (enthusiastic or neutral) and a sequence condition (massed or spaced), such that condition assignments were unknown to the research assistant. Participants clicked to provide informed consent on the first screen, and then the study began with a video of the instructor introducing herself, welcoming them to the experiment and explaining the task, either in an enthusiastic or a neutral manner. Although the delivery of the instructions differed between these two conditions, both introductory video clips informed participants that they would study pairs of Indonesian and English words, that their goal was to learn the translation of each Indonesian word, and that they would receive a monetary bonus of \$0.25 for each Indonesian word they correctly translated during the second session. Participants could not control video playback, all videos were presented exactly once, each in its entirety.

Following these initial instructions, participants then studied a total of 144 virtual flashcards (36 word pairs, ordered randomly, repeated 4 times each), each with an Indonesian word on the front and the corresponding English translation on the back. Participants clicked the flashcard to flip it and reveal the English translation. Participants were only able to flip each card

once, but were free to spend as much time as they wished on each side of the flashcard (with the constraint that the entire study phase could not last more than one hour). Participants in the massed condition saw the 4 repetitions of each word pair back-to-back, whereas the other half of participants (spaced condition) saw each repetition only after studying all other word pairs.

A short break was included after card 72 (halfway through the 144 trials). During the break, participants watched a brief video informing them that it was the halfway point and reminding them to take their time and stay focused. This video was either enthusiastic or neutral, matching the participant's initial condition assignment.

After participants had studied all flashcards, another video (either enthusiastic or neutral) asked them to complete a short demographic survey, and a final video (again, either enthusiastic or neutral) thanked them and asked them to confirm their next appointment with the research assistant. These extra procedural steps were intended to reduce the likelihood that participants would write down any word pairs immediately following the study phase.

Upon returning to the lab a week later, an undergraduate research assistant directed each participant to a computer where the recall task would take place; whenever possible we used the same cubicle that the participant used in the study session. A short video (either enthusiastic or neutral, again according to the participants' initial condition assignments) welcomed them back to the experiment and described the recall task. All participants saw the 36 Indonesian words in random order, one at a time, in the center of the screen and had to type the English translations. The recall task was self-paced and participants were allowed to leave the field empty if they did not remember the translation, however, participants could not go back to change previously-submitted responses. Slight misspellings and near-synonyms were tolerated for participant payment bonuses, but only exact matches with the previously-studied English translation were counted as correct for data analysis. After the recall task, participants completed a short questionnaire, overall accuracy was then displayed on the screen, and the performance bonus was calculated and processed. This final questionnaire included three items adapted from the enthusiasm scale of a standardized student evaluation instrument (SEEQ; [56]), "Ashley was enthusiastic about this activity," "Ashley was dynamic and energetic in conducting this activity," and "Ashley's style of presentation held my interest" (rated on a 5-point scale of agreement). We also included a question about the memorability of the instructor ("I have strong memories of the video of Ashley"), and participants also self-assessed their own task motivation ("I felt motivated to do well in this experiment").

Bayesian data analysis. We used Bayesian estimation methods for all our data analyses, which has different analytical goals than null hypothesis significance testing [57]. Rather than controlling long-run false-positive rates (as in null hypothesis significance testing), this analysis produces an informative posterior distribution of the most credible descriptions of the data. A major advantage of this approach is that the analysis directly describes an estimate of uncertainty, while avoiding conceptual problems with frequentist alternatives like confidence intervals [58, 59]. Bayesian estimation involves specifying an analytical model of how experimental effects influence an outcome measure, then estimating credible parameter values for the model given the observed data, and ultimately investigating the distribution of credible parameter values to make inferences about experimental effects (for an accessible introduction to Bayesian estimation, see [60]).

Our analysis model for test performance assumes that the number of correct responses, y , for a given participant, p , is generated by a binomial distribution with probability of success, θ_p , out of 36 English-Indonesian test item trials.

$$y_p \sim \text{Binomial}(\theta_p, 36)$$

We assume that θ values for each participant are drawn from a beta distribution that is specific to participants who had the same kind of instruction (enthusiastic or neutral, e) and sequence of study items (massed or spaced, s). The beta distribution's shape parameters, α and β , were parameterized in terms of the mode (μ) and concentration (κ) of the distribution [59].

$$\theta_p \sim \text{Beta}(\alpha_{e,s}, \beta_{e,s})$$

$$\alpha_{e,s} = \mu_{e,s} \times (\kappa_{e,s} - 2) + 1$$

$$\beta_{e,s} = (1 - \mu_{e,s}) \times (\kappa_{e,s} - 2) + 1$$

The parameters μ and κ (the mode and concentration of item response, respectively) are our primary interest in describing the group-level effects of enthusiasm and study sequence. The parameter μ is a direct estimate of overall test performance (out of 36 trials) in an enthusiasm/sequence condition (denoted by the subscripts e and s), and κ is an estimate of the group-level variance of test performance for that condition. We put vague priors on μ (as drawn from a uniform beta distribution) and κ (as drawn from a gamma distribution), informed only by the scale of the data.

$$\mu_{e,s} \sim \text{Beta}(1, 1)$$

$$\kappa_{e,s} \sim \text{Gamma}(\text{mode} = 10, \text{sd} = 100)$$

For the analysis of questionnaire responses (participant's subjective ratings of the instructor's enthusiasm and of their own motivation), we treated rating scales as numeric measures, and used a standard Bayesian hierarchical model for robust comparison of two groups [61]. In this model, the mean of both groups combined is used as the prior for each individual group mean, and the prior for each group's standard deviation is 5 times the full sample's standard deviation.

We estimated the posterior distributions of the models using JAGS [62] and the runjags package [63] for R [64]. We used four parallel Markov chain Monte Carlo (MCMC) runs; each of the four chains had 1,000 steps of adaptation and 4,000 steps of burn-in before sampling 60,000 times from the posterior distribution. The final sample was thinned to every three steps, for a total sample of 80,000 (4 chains x 20,000 samples). The effective sample size for the parameters of interest were no less than 20,000, well above the 10,000 recommended by Kruschke [59]. The Gelman-Rubin convergence diagnostic [65] was 1.0 for all parameters, indicating proper convergence of the MCMC chains.

The posterior is a probability distribution of credible parameter estimates for the model, given the observed data and the priors. A posterior distribution's 95% highest density interval (HDI) defines the boundaries of the most credible estimates of the parameter, so that the values inside the HDI are more probable than those outside. When a parameter's 95% HDI overlaps zero, it remains reasonably probable that the parameter is zero, or that the estimate is so close to zero that it is not theoretically meaningful.

Results

Participants in the enthusiasm condition rated the simulated instructor, Ashley, as marginally more enthusiastic ($M = 3.4$, $SD = 0.5$; on a zero-to-4-point scale, the mean of the 3 SEEQ items) than participants in the neutral condition ($M = 3.2$, $SD = 0.6$), but the 95% HDI of the posterior distribution of this difference of means included zero and values close to zero

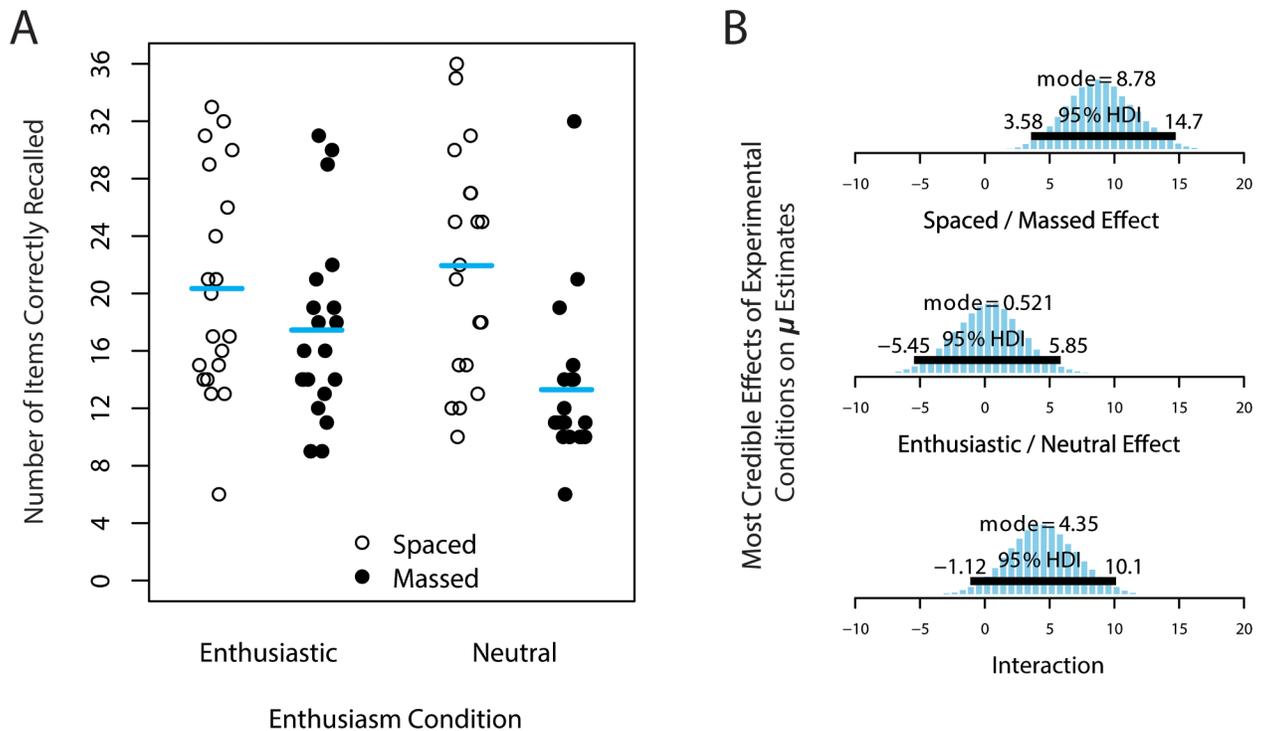


Fig 2. Number of items correctly recalled in Experiment 1. (A) Number of items correctly recalled (out of 36 possible) in Experiment 1. Each participant's raw score is a point; horizontal bars mark the arithmetic mean of correct recall in each condition. (B) Posterior estimates of experimental effects on μ , the central tendency of item recall, for (top) spaced—massed, (middle) enthusiastic—neutral, and (bottom) the interaction between these two factors (benefit of enthusiasm in massed—benefit of enthusiasm in spaced). The 95% highest density interval (HDI) shows the range of values that contains 95% of the posterior distribution, so that the values inside the HDI are more probable than those outside.

<https://doi.org/10.1371/journal.pone.0181775.g002>

(mode = 0.203; 95% HDI: -0.0474 to 0.501). On another item, “I have strong memories of Ashley,” participants in the enthusiasm condition also provided marginally-higher ratings ($M = 3.0$, $SD = 0.8$; on a 1-to-5-point scale) than the neutral condition ($M = 2.6$, $SD = 1.0$), but again, the posterior distribution of the difference of these means overlapped zero (mode = 0.346; 95% HDI: -0.0747 to 0.779). There was no difference between enthusiasm conditions in participants' responses to the item “I felt motivated to do well in this experiment” (both enthusiastic and neutral were $M = 2.8$, $SD = 0.82$ in enthusiasm condition, $SD = 0.72$ in neutral condition).

On average, out of 36 pairs, participants provided the correct English translation for 18.2 items (50.6%; $SD = 7.6$), provided incorrect responses on 13.5 items (37.5%; $SD = 7.7$), and left blank the remaining 4.3 items (11.9%; $SD = 6.3$). Fig 2A illustrates the number of correctly recalled items for each of the experimental conditions.

Our analysis model included group-level parameter estimates of average items correctly recalled, μ , for each of the four combination of conditions (enthusiastic/massed; enthusiastic/spaced; neutral/massed; and neutral/spaced). Collapsing across sequence conditions (spaced and massed), the difference between enthusiastic and neutral μ estimates at every step in the MCMC chains creates a posterior distribution of memory improvement due to enthusiasm. The mode of this distribution is 0.521 items (95% HDI: -5.45 to 5.85 items), suggesting that the most credible estimates of the main effect of enthusiastic instructions are close to zero. Similarly, collapsing across enthusiasm conditions, we can calculate the difference between spaced and massed parameter estimates, which yields a posterior distribution with a mode of 8.78 items (95% HDI: 3.58 to 14.7 items). From this posterior distribution, we can infer that there

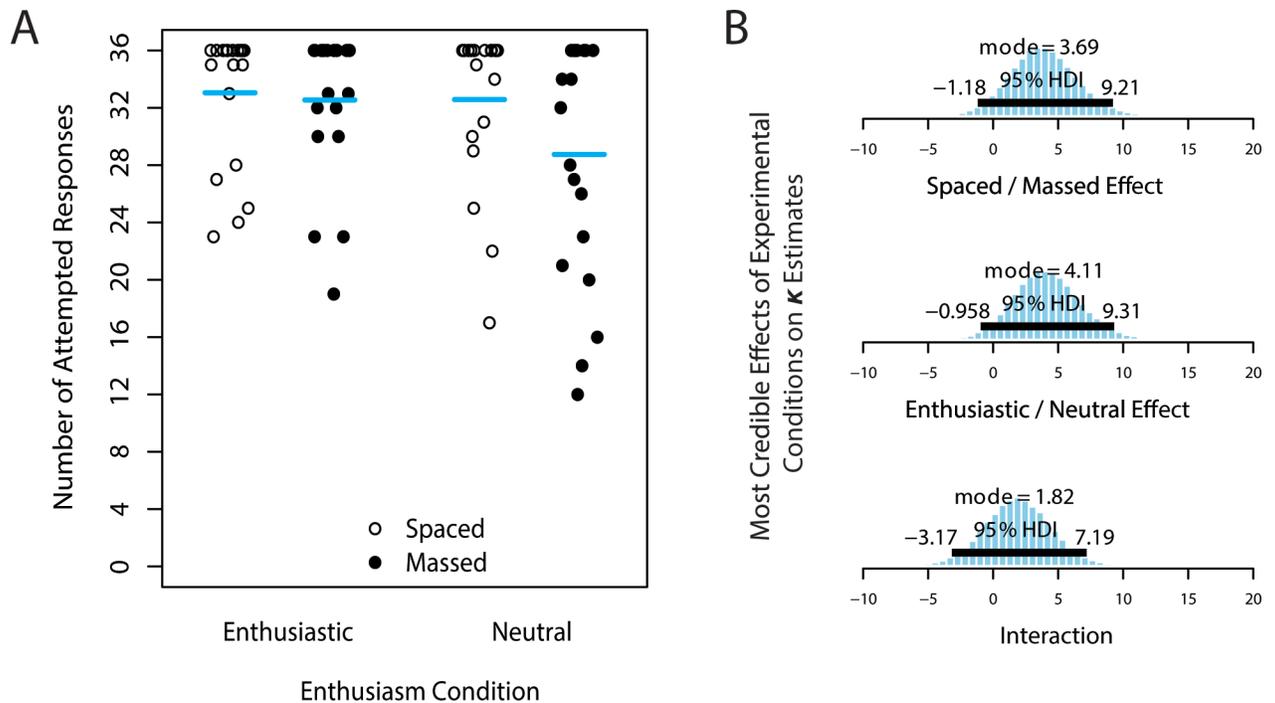


Fig 3. Number of attempted responses in Experiment 1. (A) Number of attempted (non-blank) responses (out of 36 opportunities), irrespective of response correctness. Each participant’s number of attempted responses is a point; horizontal bars mark the arithmetic mean of the dataset. The modal value in all conditions is 36 (note the high concentration of points at the top of the chart). (B) Posterior estimates of experimental effects on κ , a concentration parameter that is inversely proportional to the “spread” of attempted responses (κ values increase as the data are more concentrated near the mode of 36). Contrasts are shown for (top) spaced—massed, (middle) enthusiastic—neutral, and (bottom) the interaction between these two factors (increased concentration of enthusiasm in massed—increased concentration of enthusiasm in spaced).

<https://doi.org/10.1371/journal.pone.0181775.g003>

was a 3 to 15 item improvement when the study sequence was spaced compared with massed. Contrasting these main effects, the effect of sequencing is greater than the effect of enthusiasm (difference mode = 8.01 items; 95% HDI: 0.697 to 18.1 items). And lastly, by taking the difference of differences between these four parameters (benefit of enthusiasm in massed—benefit of enthusiasm in spaced), we can also estimate the posterior distribution of the interaction effect between these two factors, which has a mode of 4.35 items (bigger effect of enthusiasm in the massed condition than the spaced condition; 95% HDI: -1.12 to 10.1 items), but given that the 95% HDI spans zero, it is unclear whether the interaction effect is reliably different from zero, given these data. The full posterior distributions of these differences in parameter estimates are shown in Fig 2B.

The number of correctly-recalled items is an objective measure of learning, but alternative recall measures might also be sensitive to the enthusiasm manipulation. Specifically, we also analyzed total attempted responses, the number of non-blank answers provided during recall (combining both correct and incorrect responses). The majority of participants (51.9%) provided some response to all of the 36 Indonesian word prompts, but there was some variability in the “spread” of the number of attempted responses across conditions (illustrated in Fig 3A). In our analysis model (which can similarly be applied to the total number of attempts), the variance of a distribution is estimated by κ , the concentration parameter of the beta distribution, a measure of how “bunched” the observations are around the mode [59]. Given that the mode of the distribution is at the maximum value, an increase in κ indicates an increase in the number of attempted words, and naturally accounts for the ceiling effect in the current data. In

other words, considering that the mode in all four experimental conditions (parametrized by μ) is 36, only the concentration parameter (κ) would be expected to account for differences between conditions in the total number of responses (the arithmetic mean of the dataset is not estimated in this analysis model; only mode and concentration). Similar to the previous analysis of μ for correctly-recalled items, we calculated differences in κ estimates between levels of each factor, creating posterior distributions of the factor's effect on total attempted responses. These posterior distributions are shown in Fig 3B. The most credible estimate of the spacing effect on κ is 3.69 (95% HDI: -1.18 to 9.21), and for the effect of enthusiasm, κ is estimated at 4.11 (95% HDI: -.958 to 9.31). Because both posterior distributions overlap zero, neither of these trends reached our decision threshold for being reliably different from no effect, although they do raise interesting questions for further examination.

Discussion

Enthusiastic task instructions did not have any main effect on memory performance in Experiment 1. This absence of an effect of enthusiasm lies in stark contrast to the robust observed effect of study sequence (spaced vs. massed repetitions of an item) on recall, which has been well-established in the literature (for a review, see [38]). One obvious inference then, which cannot be rejected, is that enthusiastic instruction does not improve learning. However, there are other possible interpretations.

Perhaps the experimental manipulation of enthusiasm was not sufficiently strong to spur a detectable improvement. After all, participants' subjective ratings of Ashley's enthusiasm and her memorability were just slightly higher in the enthusiasm condition, insufficient to determine a nonzero difference estimate. In real classrooms, repeated and extended exposure to an enthusiastic instructor may yield a more robust effect. Within the current study design, using short video clips of a simulated instructor, should we even be able to observe such an effect of instructional enthusiasm? We believe so. Using short videos to manipulate and control instructor enthusiasm is typical in this literature (e.g., [16]), short instructional videos are common and useful tools in contemporary blended learning contexts [66], and in the current study, considering that the video was the only source of task instructions, performance strongly hinged on participants' attention to the videos.

Another possibility is that enthusiasm interacts with task difficulty. We observed a marginal interaction such that the benefit of enthusiasm was greater in the massed condition than the spaced condition. Considering that spaced practice implicitly requires greater study effort [67–69], there might be less opportunity for enthusiastic instructions to benefit learning; but for massed practice, there might be room for enthusiasm to impel students to try harder. However, the posterior distribution for this interaction effect overlapped zero and values close to zero, so this inference fails to receive clear support from the present data.

One possibility then, is that the enthusiasm manipulation had some small effect on performance, perhaps isolated to the more difficult massed condition, but the sample was not large enough to reliably detect it. Similarly, the results hinted toward a small effect of enthusiasm on total attempted responses (perhaps enthusiasm causes students to guess more often), but the most credible parameter estimates included zero and values close to zero. Conventional (frequentist) measures of power suggest that, in order to achieve an 80% likelihood of observing a significant interaction between enthusiasm and study sequence, we would need approximately twice as many participants. This is based on the assumption that the effect size of the interaction is close to what we observed in this study.

The monetary bonus that participants received for correct recall represents another possible culprit for the lack of effect of enthusiasm. Enthusiasm was predicted to improve motivation

[34], but it is possible that current participants were principally motivated by the financial incentive. Indeed, past researchers have suggested that an instructor's expressiveness will only improve learning when students have no other extrinsic incentive to achieve [24, 25], and participants' self-assessments of their own level of motivation were practically equivalent between enthusiasm conditions. We opted to use financial incentives in Experiment 1 to increase the likelihood that student participants would study hard during the first session and return to the lab for the second session, but these may have reduced the potency of enthusiasm's effect on task motivation.

In summary, there were several possible explanations for the failure to observe an effect of enthusiasm on memory performance: financial incentives outweighing the motivating effect of instructor enthusiasm, insufficient sample size to reliably detect effects, and small differences in participants' subjective ratings of the enthusiasm manipulation. Experiment 2 was designed to address these issues, testing the possibility that an effect of enthusiasm might become apparent with a larger sample size, and without extrinsic incentives (which might result in a more robust manipulation of enthusiasm). Moreover, considering that the enthusiastic instructions, in total, were 31 seconds (12.6%) longer than the neutral condition, we also aimed to equate the lengths of these videos, eliminating the possibility that length of instruction would affect task performance.

Experiment 2 (Confirmatory study)

Experiment 2 was similar to Experiment 1, but with three important differences: (1) We increased the sample size, more than doubling the number of participants; (2) We removed the monetary incentive for memory performance; and (3) We produced new videos that better-equated the lengths of enthusiastic and neutral instructions, and to generalize Experiment 1's results to a new instructor. We preregistered the study on OSF, with the plan to use the same analysis model used for the proportion of correct responses in Experiment 1 with a larger sample size (recorded at <https://osf.io/qmg95/>). All materials and data for Experiment 2 are available at <http://osf.io/j2h8y/>.

Method

Ethics statement. The experimental procedures, materials, and recruitment protocols were approved by the Indiana University Institutional Review Board. All participants provided informed consent electronically at the start of the experiment.

Participants. Two hundred and forty-one undergraduate students at Indiana University volunteered in exchange for credit toward their Introductory Psychology experiment participation requirement. We excluded 42 participants because they did not return for the second session ($N = 14$), the unique subject identifier was accidentally assigned to more than one participant ($N = 3$), the dataset was incomplete due to skipped trials or questions ($N = 14$), they spoke Indonesian ($N = 3$), or they were enrolled in Introductory Psychology with Benjamin Motz ($N = 8$; one of the authors of this study, and the simulated instructor for Experiment 2; see *Video Recorded Instructions*, below). After exclusions there were 199 participants, 111 females (by self-report; 55.8%), with an average age of 18.9 years ($SD = 1.2$). One hundred sixty nine of these participants (85%) reported being native English speakers, and the remaining non-native English speaking participants had been speaking English for an average of 11.1 years ($SD = 4.8$).

Stimuli. We used the same English-Indonesian word pairs as Experiment 1, but to improve the sensitivity of this stimulus set, word pairs that were successfully recalled by at least 75% of participants in Experiment 1 were replaced with new pairs from [54]. There were 9

pairs from Experiment 1 that met this criterion, each close cognates (e.g., Photocopy / Fotokopi). We selected 9 replacement pairs that avoided any such surface similarity between the English and Indonesian tokens.

Video recorded instructions. We edited the scripts used in Experiment 1 so that the lengths of the enthusiastic and neutral scripts were evenly matched, and to remove reference to any financial bonuses for correct recall. Instead of extrinsic incentives, both scripts simply asked participants to invest effort in the task because it would benefit science, and because the researchers would appreciate it.

With these new scripts came the opportunity to recreate the videos with a new instructor, seeking further generalizability. Rather than recruiting another actor, one of the authors of this article (Motz), an experienced college teacher, memorized the scripts and recorded himself delivering the instructions, enthusiastic and neutral, in the same manner as described in Experiment 1. These videos were recorded in a small classroom in the Psychology Building (see Fig 1 for representative stills; the individuals shown in this figure have given written consent to publish these images).

The total duration of all recorded video segments in the enthusiastic condition was 5 minutes 31 seconds, and 5 minutes 15 seconds in the neutral condition. This small difference in duration is caused by expressive pauses in the enthusiastic videos, not because the enthusiastic videos had more content or information.

Procedure. The procedure for Experiment 2 was identical to Experiment 1, except that participants did not receive cash bonuses for performance, and the simulated instructor's name was updated in questionnaire items.

Bayesian data analysis. The analysis models and method for Experiment 2 were identical to Experiment 1. To provide the most conservative examination of experimental effects, we used the same vague priors as Experiment 1.

Results

The enthusiasm manipulation had a stronger effect on participants' ratings in Experiment 2 than in Experiment 1. Participants in the enthusiasm condition clearly rated the simulated instructor, Ben, as being more enthusiastic ($M = 3.5$, $SD = 0.65$; neutral, $M = 2.8$, $SD = 0.86$; difference mode = 0.819; 95% HDI: 0.627 to 1.01) and more memorable ($M = 3.2$, $SD = 0.93$; neutral, $M = 2.5$, $SD = 0.95$; difference mode = 0.717; 95% HDI: 0.448 to 0.983). There was also a directional trend for participants to report feeling slightly more motivated to perform well in the enthusiastic condition ($M = 3.0$, $SD = 0.80$) than the neutral condition ($M = 2.8$, $SD = 0.69$), perhaps due to the removal of the cash bonuses for performance, although credible estimates of this group difference overlapped zero (mode = 0.164; 95% HDI: -0.0408 to 0.372).

Also likely stemming from the lack of financial incentives, as well as the removal of the most frequently-recalled items, recall accuracy was lower in Experiment 2 than Experiment 1. Across all conditions, participants correctly recalled only 9.8 items (27.2%; $SD = 7.6$), provided incorrect responses on 20.7 items (57.6%; $SD = 10.1$), and left the remaining 5.5 items empty (15.2%; $SD = 8.3$). Fig 4A illustrates recall accuracy for the four different conditions.

In spite of the increased sample size (with accordingly narrower HDIs), and participants' stronger subjective ratings of enthusiasm, the most credible estimates of the effect of enthusiasm on recall accuracy still overlapped zero (mode = 1.6; 95% HDI: -1.49 to 5.37). Although this posterior distribution is shifted slightly in the direction of a very small positive benefit of instructional enthusiasm, this benefit was not reliably different from zero effect. In contrast, study sequence again had a robust main effect, our best estimate is that participants who were assigned a spaced sequence correctly recalled 7.36 more items than those who were assigned a

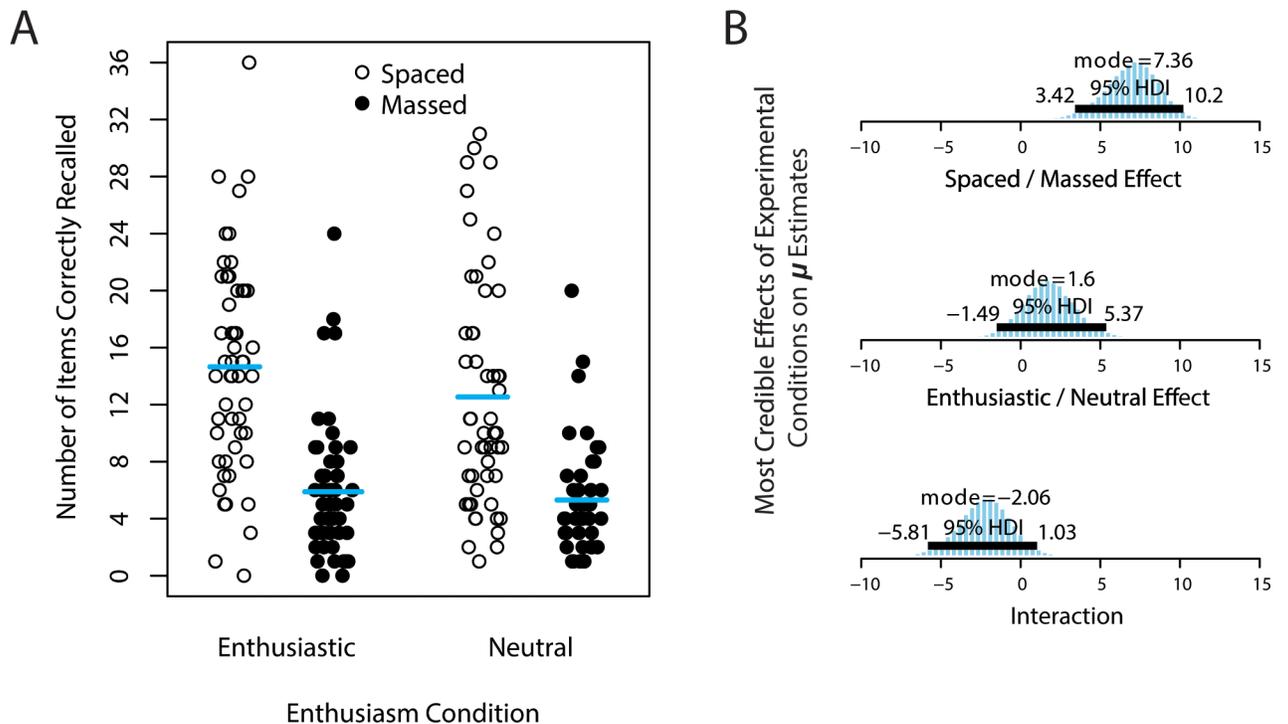


Fig 4. Number of items correctly recalled in Experiment 2. (A) Number of items correctly recalled (out of 36 possible) in Experiment 2. Each participant's score is a point; horizontal bars mark the arithmetic mean of correct recall in each condition. (B) Posterior estimates of experimental effects on μ , the central tendency of item recall, for (top) spaced—massed, (middle) enthusiastic—neutral, and (bottom) the interaction between these two factors (benefit of enthusiasm in massed—benefit of enthusiasm in spaced). The 95% highest density interval (HDI) shows the range of values that contains 95% of the posterior distribution, so that the values inside the HDI are more probable than those outside.

<https://doi.org/10.1371/journal.pone.0181775.g004>

massed sequence (95% HDI: 3.42 to 10.2). Although we observe a large and reliable effect of study sequence, the slight positive shift in the effect of enthusiasm meant that the difference between these main effects (sequence minus enthusiasm) overlapped zero (difference mode = 5.51 items; 95% HDI: -0.305 to 10.0). Also, the posterior distribution of the interaction between sequence and enthusiasm again spanned zero (mode = -2.06; 95% HDI: -5.81 to 1.03), and if anything, the most credible estimates suggest a reversal from Experiment 1 (improvement due to enthusiasm is now smaller in the massed condition than the spaced condition).

Although participants had lower overall accuracy in Experiment 2, the majority (50.3%) still provided a response to all 36 of the Indonesian word prompts. The remaining 49.7% of participants left at least one item blank. Fig 5A illustrates the number of attempted responses across all conditions. We conducted an exploratory analysis on the likelihood of a participant providing an answer, regardless of accuracy. This is measured by κ , the concentration of attempts near the maximum of 36 items. Posterior estimates of κ suggest that enthusiasm increased the likelihood of providing a response (mode of κ differences = 2.39; 95% HDI: 0.82 to 4.02), but κ was not reliably affected by study sequence (mode = 0.218; 95% HDI: -1.34 to 1.85), nor was there an interaction between study sequence and enthusiasm (mode = 0.589; 95% HDI: -1.04 to 2.16).

Discussion

Experiment 2 was a confirmatory study, preregistered to examine the hypothesis that an effect of instructional enthusiasm on memory performance, or an interaction between enthusiasm

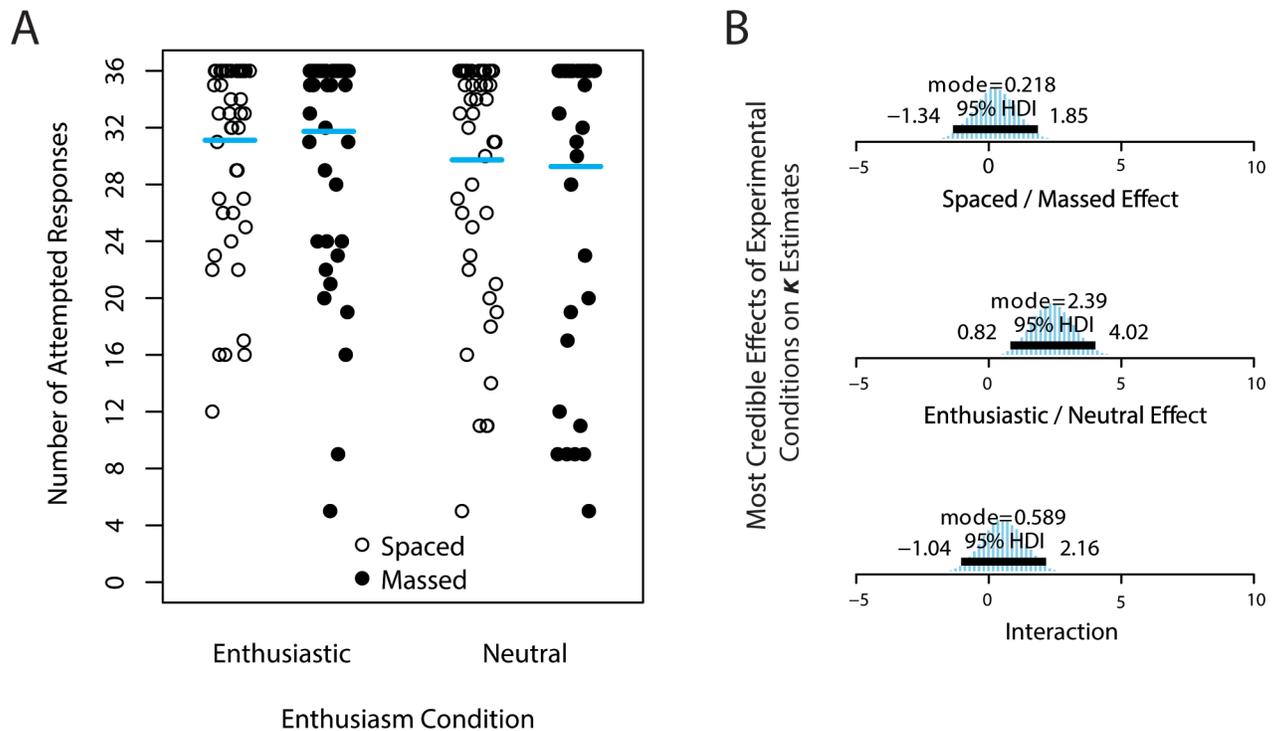


Fig 5. Number of attempted responses in Experiment 2. (A) Number of attempted (non-blank) responses (out of a maximum 36), irrespective of response correctness, in Experiment 2. Each participant's number of attempted responses is a point; horizontal bars mark the arithmetic mean of the dataset. The modal value in all conditions is 36 (note the high concentration of points at the top of the chart). (B) Posterior estimates of experimental effects on κ , a concentration parameter that is inversely proportionate to the "spread" of attempted responses (κ values increase as the data are more concentrated near the mode of 36). Contrasts are shown for (top) spaced—massed, (middle) enthusiastic—neutral, and (bottom) the interaction between these two factors (increased concentration of enthusiasm in massed—increased concentration of enthusiasm in spaced).

<https://doi.org/10.1371/journal.pone.0181775.g005>

and study sequence, would become apparent with a larger sample. These hypotheses were not supported.

We failed to observe any reliable effect of enthusiastic instructions on learning performance despite a large sample size, and also despite manipulation checks confirming that the intervention had a strong effect on participants' subjective ratings of the instructor's enthusiasm and the memorability of the instructions. And enthusiasm was not psychologically inert. Although there was no effect on recall accuracy, our exploratory analysis found that enthusiastic task instructions caused participants to provide more answers to the Indonesian word prompts (or, in other words, they were less likely to leave items blank), as we observed the distribution of items answered to be more concentrated around the maximum value of 36 in the enthusiasm conditions. We observed this same directional pattern in Experiment 1 (although the 95% HDI overlapped zero), but in Experiment 2, with increased power, enthusiastic instructions reliably increased the likelihood of providing a response. There was no penalty for incorrect responses in the current study, and since there was no difference in the number of correctly-recalled items, the additional responses in the enthusiasm condition were incorrect (many were either incorrect cognates, such as 'liburan' translated as 'librarian' instead of 'holiday,' or words that would be correct translations on other items).

As an alternative to merely leaving an item blank, an increased tendency to provide a response, even if incorrect, is evidence of a more liberal response criterion [70]. In classic memory research, experimental manipulations of task motivation have been shown to affect

this response criterion, resulting in increased response attempts even in the absence of improvements in recall accuracy. Explicit instructions and encouragement to provide more responses [53, 71, 72] and financial incentives [52] may increase the number of responses, but past studies have shown these are not typically accompanied by increased quality of the memory trace or accuracy of the responses [70, 73]. It would seem that enthusiastic instruction has an effect that is analogous to encouraging increased responses, incentivizing participants to be more engaged with the task, less likely to leave items blank, but not providing any measurable advantage for learning or memory.

General discussion

Based on the results of the current study, we conclude that if there is any isolatable, direct effect of instructional enthusiasm on recall accuracy, it is small and unreliable. Across two carefully controlled experiments, we found that participants exposed to an enthusiastic instructor do not demonstrate reliable improvements in how much they learn during a memory task.

Since Rosenshine's [31] early claims of "strikingly consistent results" (p. 506) relating enthusiasm to pupil achievement, there have been two divergent trends in the literature. On the one hand, experimental research has severely disrupted this consistency—null effects of enthusiastic instructions on test performance, such as those reported in the current study, have steadily mounted [20–24, 27–29, 74]. On the other hand, students and educators have both become more steadfast in their endorsement of enthusiasm as a crucial aspect of effective teaching [12].

How can we reconcile enthusiasm's broad popularity in education with our failure to observe a consistent benefit of an instructor's enthusiasm on learning? Perhaps the answer lies in accepting that motivated task behavior does not necessarily lead to memory improvement.

A dissociation between learning and motivation

Stated broadly, the current results suggest that enthusiasm, of the kind manipulated here, has little practical value for memory performance. If the goal of an academic exercise is to help a student remember the to-be-learned material, enthusiastic instructions may not have direct benefit toward this goal. Our results suggest that learners would probably reap better outcomes with an optimized studying sequence (or perhaps other cognitive strategies [47]) than by exposure to an expressive and dynamic instructor. Lest these conclusions should seem surprising to teaching enthusiasts, recent research has gone even further to demonstrate that an instructor's fluency (the ease with which students perceive to-be-learned material) has no direct effect on learning [75, 76].

Even though instructor enthusiasm has no effect on recall accuracy, results from Experiment 2 indicate that it still had motivational effects on task behavior, increasing self-reported interest in the instructor's presentation, and increasing the number of responses provided during test. Our results support and extend those of Bettencourt et al. [21], who also observed evidence of increased student motivation toward task behavior, but no effect on test performance, due to an otherwise robust experimental manipulation of instructor enthusiasm. Patrick et al. [34] similarly found that enthusiastic instructions had a contagious effect on students' self-reported levels of motivation during a learning activity, but there was no effect on a behavioral measure of self-directed learning. Rather than accept a dissociation between motivation and achievement, these past researchers attributed their mixed results to methodological nuance. The folk theory that enthusiastic teaching begets student motivation, increasing interest and attention, which then improves learning [31], was too compelling to dispute [21, 51]. However research in cognitive psychology has repeatedly shown that learning performance does not necessarily benefit from increased interest [77, 78] or increased motivation [52, 79].

The current results add to mounting evidence indicating that a motivated student is not necessarily a *learned* student—a fact that may intuitively resonate with teachers who have encountered students who try hard, but seem unable to accurately recall the material. Motivation alone does not improve test performance. Rather, as observed in the current study, motivation may instead manifest as increased self-reported interest in an activity, and as being more likely to respond to explicit prompts. These are nontrivial outcomes, but recall accuracy does not necessarily hinge on them.

Implications

The present lack of an influence of instructor enthusiasm on memory accuracy may, of course, be restricted to either our manipulation of enthusiasm, or to the learning task. Prolonged, chronic instructor enthusiasm may have an influence on performance, may affect a more meaningful learning task, or may have isolated effects on students depending on their motivations and goals. Nonetheless, short “doses” of instructor enthusiasm (such as those commonly manipulated in this line of research) are likely to be of increasing practical relevance in contemporary educational environments where the instructor is not directly supplying learning materials, but rather is directing students’ active learning efforts (discussed in [34]). It is noteworthy, then, that we observed divergent effects when concurrently manipulating instructional behaviors and learning materials; the instructor’s enthusiasm did have positive influences on learners’ test effort and self-reports of engagement, but only the properties of the learning materials (item sequence) reliably affected test performance.

Enthusiasm’s relationship to student achievement (commonly operationalized as test performance) is not supported. However, enthusiasm is not without virtue (unless the reader believes that the ultimate goal of education is merely to improve test performance). Its effect on student motivation may serve important practical purposes in the classroom, such as seeding interest in a topic, or engaging the learner in an activity. Just because it does not significantly improve recall of information does not imply that educators should jettison enthusiasm from schools, from teaching evaluation instruments, or from the advice given to new instructors. Claims regarding enthusiastic teaching’s benefits should, however, be limited to improvements in task motivation and enjoyment (which are well-supported by empirical research, including the current study); and such claims should not extend to improvements in test performance. Education is a fundamentally social activity, and teachers’ behaviors have important consequences beyond memorization [80–83]. This complexity is inherent in educational systems, and applications of memory research will benefit from future experimental examinations of the relationship between instructional techniques and student learning [84].

Acknowledgments

This study benefited greatly from Ashley Dillard’s chameleonic performances of the instructional scripts in Experiment 1, as well as Dustin Finch, Chirag Divecha, Joshua Noll, and Mo Nguyen’s assistance with data collection.

Author Contributions

Conceptualization: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho, Robert L. Goldstone.

Data curation: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho, Kaley L. Liang.

Formal analysis: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho.

Funding acquisition: Robert L. Goldstone.

Investigation: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho, Kaley L. Liang.

Methodology: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho, Robert L. Goldstone.

Project administration: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho, Kaley L. Liang, Robert L. Goldstone.

Resources: Benjamin A. Motz, Joshua R. de Leeuw, Kaley L. Liang.

Software: Joshua R. de Leeuw.

Supervision: Benjamin A. Motz, Paulo F. Carvalho, Kaley L. Liang, Robert L. Goldstone.

Validation: Joshua R. de Leeuw, Paulo F. Carvalho.

Visualization: Paulo F. Carvalho.

Writing – original draft: Benjamin A. Motz.

Writing – review & editing: Benjamin A. Motz, Joshua R. de Leeuw, Paulo F. Carvalho, Kaley L. Liang, Robert L. Goldstone.

References

1. Lowman J. Mastering the techniques of teaching. 2nd ed. San Francisco: Jossey-Bass Publishers; 1995.
2. Zimbardo PG. A passion for psychology: Teaching it charismatically, integrating teaching and research synergistically, and writing about it engagingly. In: Zimbardo PG, Sternberg RJ, editors. *Teaching Introductory Psychology: Survival tips from the experts*. Washington, DC: American Psychological Association; 1997. p. 7–34.
3. Davis BG. *Tools for teaching*. 2nd ed. San Francisco, CA: Jossey-Bass; 2009.
4. Barr AS. *Characteristic differences in the teaching performance of good and poor teachers of the social studies*. Bloomington, Illinois: Public School Publishing Company; 1929.
5. McKeachie WJ, Svinicki MD. *Teaching tips: Strategies, research, and theory for college and university teachers*. 13th ed. Boston: Cengage; 2010.
6. Feldman KA. The superior college teacher from the students' view. *Research in Higher Education*.
7. Senko C, Belmonte K, Yakhkind A. How students' achievement goals shape their beliefs about effective teaching: A build-a-professor' study. *Brit J Educ Psychol*. 2012; 82(3):420–35. <https://doi.org/10.1111/j.2044-8279.2011.02036.x8>
8. Marsh HW. Multidimensional students evaluations of teaching effectiveness: A test of alternative higher-order structures. *J Educ Psychol*. 1991; 83(2):285–96. <https://doi.org/10.1037/0022-0663.83.2.2859>
9. Marsh HW. Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student course instructor characteristics. *J Educ Psychol*. 1983; 75(1):150–66. <https://doi.org/10.1037/0022-0663.75.1.150>
10. Isaacson RL, McKeachie WJ, Milholland JE. Correlation of teacher personality variables and student ratings. *J Educ Psychol*. 1963; 54(2):110–7. <https://doi.org/10.1037/h0048797>
11. Murray HG, Rushton JP, Paunonen SV. Teacher personality traits and student instructional ratings in six types of university courses. *J Educ Psychol*. 1990; 82(2):250–61.
12. Keller MM, Hoy AW, Goetz T, Frenzel AC. Teacher Enthusiasm: Reviewing and Redefining a Complex Construct. *Educ Psychol Rev*. 2016; 28(4):743–69. <https://doi.org/10.1007/s10648-015-9354-y>
13. Frenzel AC, Goetz T, Pekrun R, Watt HMG. Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*. 2010; 20(2):507–37. <https://doi.org/10.1111/j.1532-7795.2010.00645.x>
14. Mastin VE. Teacher enthusiasm. *J Educ Res*. 1963; 56(7):385–6.
15. Coats WD, Smidchen U. Audience recall as a function of speaker dynamism. *J Educ Psychol*. 1966; 57(4):189–91. <https://doi.org/10.1037/h0023568>

16. Ware JE Jr., Williams RG. The Dr. Fox effect: a study of lecturer effectiveness and ratings of instruction. *J Med Educ.* 1975; 50(2):149–56. PMID: [1120118](https://pubmed.ncbi.nlm.nih.gov/1120118/)
17. Carlisle C, Phillips DA. The effects of enthusiasm training on selected teacher and student behaviors in preservice physical education teachers. *Journal of Teaching in Physical Education.* 1984; 4(1):64–75. <https://doi.org/10.1123/jtpe.4.1.64>
18. Streeter BB. The effects of training experienced teachers in enthusiasm on students' attitudes toward reading. *Reading Psychology.* 2006; 7(4):249–59. <https://doi.org/10.1080/0270271860070403>
19. Brigham FJ. Teacher enthusiasm in learning disabilities classrooms: Effects on learning and behavior. *Learning Disabilities Research and Practice.* 1992; 7(2):68–73.
20. Larkins AG, McKinney CW. Two studies of the effects of teacher enthusiasm on the social studies achievement of seventh grade students. *Theory & Research in Social Education.* 1982; 10(1):27–41. <https://doi.org/10.1080/00933104.1982.10505417>
21. Bettencourt EM, Gillett MH, Gall MD, Hull RE. Effects of teacher enthusiasm training on student on-task behavior and achievement. *Am Educ Res J.* 1983; 20(3):435–50. <https://doi.org/10.3102/00028312020003435>
22. McKinney CW, Larkins AG, Kazelskis R, Ford MJ, Allen JA, Davis JC. Some effects of teacher enthusiasm on student achievement in fourth grade social studies. *J Educ Res.* 1983; 76(4):249–53.
23. Stewart RA. Interaction effects of teacher enthusiasm and student notetaking on recall and recognition of lecture content. *Communication Research Reports.* 1989; 6(2):84–9. <https://doi.org/10.1080/08824098909359838>
24. Williams RG, Ware JE. Validity of student ratings of instruction under different incentive conditions: A further study of the Dr. Fox effect. *J Educ Psychol.* 1976; 68(1):48–56.
25. Marsh HW, Ware JE. Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New interpretations of the Dr Fox Effect. *J Educ Psychol.* 1982; 74(1):126–34. <https://doi.org/10.1037/0022-0663.74.1.126>
26. Evertson CM, Anderson CW, Anderson LM, Brophy JE. Relationships between classroom behaviors and student outcomes in junior-high mathematics and english-classes. *Am Educ Res J.* 1980; 17(1):43–60. <https://doi.org/10.3102/00028312017001043>
27. Perry RP, Abrami PC, Leventhal L. Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. *J Educ Psychol.* 1979; 71(1):107–16. <https://doi.org/10.3102/00346543052003446>
28. Andersen JF, Withrow JG. The impact of lecturer nonverbal expressiveness on improving mediated instruction. *Commun Educ.* 1981; 30(4):342–53. <https://doi.org/10.1080/03634528109378490>
29. Wood AM. The effects of teacher enthusiasm on student motivation, selective attention, and text memory. PhD dissertation, University of Western Ontario. 1998.
30. Collins ML. Effects of enthusiasm training on preservice elementary teachers. *J Teach Educ.* 1978; 29(1):53–7. <https://doi.org/10.1177/002248717802900120>
31. Rosenshine B. Enthusiastic teaching: A research review. *The School Review.* 1970; 78(4):499–514. <https://doi.org/10.1086/442929>
32. Murray HG. Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *J Educ Psychol.* 1983; 75(1):138–49.
33. Kunter M, Frenzel A, Nagy G, Baumert J, Pekrun R. Teacher enthusiasm: Dimensionality and context specificity. *Contemp Educ Psychol.* 2011; 36(4):289–301. <https://doi.org/10.1016/j.cedpsych.2011.07.001>
34. Patrick BC, Hisley J, Kempler T. "What's everybody so excited about?": The effects of teacher enthusiasm on student intrinsic motivation and vitality. *J Exp Educ.* 2000; 68(3):217–36. <https://doi.org/10.1080/00220970009600093>
35. Perry RP, Magnusson JL. Effective instruction and students perceptions of control in the college classroom: Multiple-Lectures Effects. *J Educ Psychol.* 1987; 79(4):453–60. <https://doi.org/10.1037/0022-0663.79.4.453>
36. Perry RP, Magnusson JL. Causal attributions and perceived performance: Consequences for college students' achievement and perceived control in different instructional conditions. *J Educ Psychol.* 1989; 81(2):164–72.
37. Pashler H, Bain PM, Bottge BA, Graesser A, Koedinger K, McDaniel M, et al. Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007–2004. National Center for Education Research; 2007.
38. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol Bull.* 2006; 132(3):354–80. <https://doi.org/10.1037/0033-2909.132.3.354> PMID: [16719566](https://pubmed.ncbi.nlm.nih.gov/16719566/)

39. Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychol Sci Publ Int.* 2013; 14(1):4–58. <https://doi.org/10.1177/1529100612453266> PMID: 26173288
40. Hatala RM, Brooks LR, Norman GR. Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Adv Health Sci Educ.* 2003; 8(1):17–26. <https://doi.org/10.1023/A:1022687404380>
41. Olina Z, Reiser R, Huang XX, Lim J, Park S. Problem format and presentation sequence: Effects on learning and mental effort among US high school students. *Appl Cognitive Psych.* 2006; 20(3):299–309. <https://doi.org/10.1002/acp.1246>
42. Sobel HS, Cepeda NJ, Kapler IV. Spacing effects in real-world classroom vocabulary learning. *Appl Cognitive Psych.* 2011; 25(5):763–7. <https://doi.org/10.1002/acp.1747>
43. Seabrook R, Brown GDA, Solity JE. Distributed and massed practice: From laboratory to classroom. *Appl Cognitive Psych.* 2005; 19(1):107–22. <https://doi.org/10.1002/acp.1066>
44. Ridgeway K, Mozer MC, Bowles AR. Forgetting of foreign-language skills: A corpus-based analysis of online tutoring software. *Cognitive Sci.* 2017; 41(4):924–49. <https://doi.org/10.1111/cogs.12385> PMID: 27344981
45. Lindsey RV, Shroyer JD, Pashler H, Mozer MC. Improving students' long-term knowledge retention through personalized review. *Psychol Sci.* 2014; 25(3):639–47. <https://doi.org/10.1177/0956797613504302> PMID: 24444515
46. Laski EV, Reeves TD, Ganley CM, Mitchell R. Mathematics teacher educators' perceptions and use of cognitive research. *Mind Brain Educ.* 2013; 7(1):63–74. <https://doi.org/10.1111/mbe.12009>
47. Rohrer D, Pashler H. Recent research on human learning challenges conventional instructional strategies. *Educational Researcher.* 2010; 39(5):406–12. <https://doi.org/10.3102/0013189x10374770>
48. Morehead K, Rhodes MG, DeLozier S. Instructor and student knowledge of study strategies. *Memory.* 2016; 24(2):257–71. <https://doi.org/10.1080/09658211.2014.1001992> PMID: 25625188
49. Bronfenbrenner U. The experimental ecology of education 1. *Educational Researcher.* 1976; 5(9):5–15.
50. Newman D, Cole M. Can scientific research from the laboratory be of any use to teachers? *Theory Into Practice.* 2004; 43(4):260–267.
51. Keller MM, Neumann K, Fischer HE. Teacher enthusiasm and student learning. In: Hattie J, Anderman E, editors. *International guide to student achievement.* New York: Routledge; 2013. p. 247–9.
52. Nilsson LG. Motivated Memory: Dissociation between performance data and subjective reports. *Psychol Res.* 1987; 49(2–3):183–8. <https://doi.org/10.1007/Bf00308685>
53. Britton BK, Meyer BJ, Hodge MH, Glynn SM. Effects of the organization of text on memory: Tests of retrieval and response criterion hypotheses. *J Exp Psychol-Hum L.* 1980; 6(5):620–9. <https://doi.org/10.1037/0278-7393.6.5.620>
54. Kornell N, Son LK. Learners' choices and beliefs about self-testing. *Memory.* 2009; 17(5):493–501. <https://doi.org/10.1080/09658210902832915> PMID: 19468957
55. de Leeuw JR. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behav Res Methods.* 2015; 47(1):1–12. <https://doi.org/10.3758/s13428-014-0458-y> PMID: 24683129
56. Marsh HW. SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of teaching. *British J Educ Psychol.* 1982; 52(1): 77–95. <https://doi.org/10.1111/j.2044-8279.1982.tb02505.x>
57. Kruschke JK, Liddell TM. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev.* 2017. <https://doi.org/10.3758/s13423-016-1221-4> PMID: 28176294
58. Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers EJ. The fallacy of placing confidence in confidence intervals. *Psychon Bull Rev.* 2016; 23(1):103–23. <https://doi.org/10.3758/s13423-015-0947-8> PMID: 26450628
59. Kruschke JK. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* 2 ed: Academic Press; 2014.
60. Kruschke JK, Liddell TM. Bayesian data analysis for newcomers. *Psychon Bull Rev.* 2017. <https://doi.org/10.3758/s13423-017-1272-1> PMID: 28405907
61. Kruschke JK. Bayesian estimation supersedes the *t*-test. *J Exp Psychol Gen.* 2013; 142(2):573–603. <https://doi.org/10.1037/a0029146> PMID: 22774788
62. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing;* 2003.
63. Denwood MJ. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *J Stat Softw.* 2016; 71(9):1–25. <https://doi.org/10.18637/jss.v071.i09>

64. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
65. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science*. 1992; 7(4):457–72. <https://doi.org/10.1214/ss/1177011136>
66. Tiernan P. An inquiry into the current and future uses of digital video in university teaching. *Education and Information Technologies*. 2015; 20(1):75–90. <https://doi.org/10.1007/s10639-013-9266-8>
67. Dempster FN. Spacing effects and their implications for theory and practice. *Educ Psychol Rev*. 1989; 1(4):309–30. <https://doi.org/10.1007/Bf01320097>
68. Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *J Mem Lang*. 2009; 60(4):437–47. <https://doi.org/10.1016/j.jml.2009.01.004>
69. Rohrer D. Avoidance of overlearning characterises the spacing effect. *Eur J Cogn Psychol*. 2009; 21(7):1001–12. <https://doi.org/10.1080/09541440802425954>
70. Koriat A, Goldsmith M. Monitoring and control processes in the strategic regulation of memory accuracy. *Psychol Rev*. 1996; 103(3):490–517. <https://doi.org/10.1037/0033-295x.103.3.490> PMID: 8759045
71. Bousfield WA, Rosner SR. Free vs unhibited recall. *Psychon Sci*. 1970; 20(2):75–6. <https://doi.org/10.3758/BF03335608>
72. Roediger HL, Srinivas K, Waddill P. How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *J Exp Psychol Gen*. 1989; 118(3):255–7. <https://doi.org/10.1037/0096-3445.118.3.255>
73. Weiner B. Motivation and memory. *Psychol Monogr*. 1966; 80(18):1–22. <https://doi.org/10.1037/h0093896> PMID: 5971462
74. Alsharif NZ, Qi Y. A three-year study of the impact of instructor attitude, enthusiasm, and teaching style on student learning in a medicinal chemistry course. *Am J Pharm Educ*. 2014; 78(7):132. <https://doi.org/10.5688/ajpe787132> PMID: 25258437
75. Carpenter SK, Mickes L, Rahman S, Fernandez C. The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *J Exp Psychol-Appl*. 2016; 22(2):161–72. <https://doi.org/10.1037/xap0000077> PMID: 26844368
76. Carpenter SK, Wilford MM, Kornell N, Mullaney KM. Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychon B Rev*. 2013; 20(6):1350–6. <https://doi.org/10.3758/s13423-013-0442-z> PMID: 23645413
77. Harp SF, Mayer RE. How seductive details do their damage: A theory of cognitive interest in science learning. *J Educ Psychol*. 1998; 90(3):414–34. <https://doi.org/10.1037/0022-0663.90.3.414>
78. McDaniel MA, Waddill PJ, Finstad K, Bourg T. The effects of text-based interest on attention and recall. *J Educ Psychol*. 2000; 92(3):492–502. <https://doi.org/10.1037/0022-0663.92.3.492>
79. Ngaosuvan L, Mäntylä T. Rewarded remembering: Dissociations between self-rated motivation and memory performance. *Scandinavian Journal of Psychology*. 2005; 46(4):8p. <https://doi.org/10.1111/j.1467-9450.2005.00462.x> PMID: 16014076
80. Brophy J. Teacher influences on student achievement. *Am Psychol*. 1986; 41(10):1069–77. <https://doi.org/10.1037/0003-066x.41.10.1069>
81. Umbach PD, Wawrzynski MR. Faculty do matter: The role of college faculty in student learning and engagement. *Research in Higher Education*. 2005; 46(2):153–84. <https://doi.org/10.1007/s11162-004-1598-1>
82. Shuell TJ. Toward an integrated theory of teaching and learning. *Educ Psychol*. 1993; 28(4):291–311. https://doi.org/10.1207/s15326985ep2804_1
83. Vermunt JD, Verloop N. Congruence and friction between learning and teaching. *Learn Instr*. 1999; 9(3):257–80. [https://doi.org/10.1016/S0959-4752\(98\)00028-0](https://doi.org/10.1016/S0959-4752(98)00028-0)
84. Koedinger KR, Booth JL, Klahr D. Education research. Instructional complexity and the science to constrain it. *Science*. 2013; 342(6161):935–7. <https://doi.org/10.1126/science.1238056> PMID: 24264979