they switched to testing themselves (i.e., seeing the cue and trying to think of the target, then checking the target) the majority of the time. In short, they preferred testing if they could get the answer right. For similar results, see Tullis, Finley, and Benjamin (2013).

Additional support for our hypotheses comes from our intuition that, as stated earlier, people like tests. This is why "wait, wait, don't tell me!" is both an idiom and a popular radio show; answering questions is more enjoyable than being told the answers. Anecdotal evidence also suggests that most people intuitively believe testing is a good way to learn. For example, in Quizlet, the wildly popular learning app that had over 50 million active users per month as of October 2018 (Clark, 2018), students can choose to study using a variety of modes (e.g., flashcards, learn, write, spell, test), but all of these modes involve retrieval. Students who use the app are implicitly buying in to the idea that testing is the most effective way (and the only way, in the app) to learn the material they learn on Quizlet. The intuition that testing is the appropriate way to learn is so strong that we have caught participants in laboratory experiments covering the target word in a word pair with their hand so they could test themselves, effectively turning a restudy condition into a test condition. This ruined our control condition, but it also testifies to the strong impulse students have to test themselves on material they know well.

The present experiments

Our ultimate goal was to increase learning by getting participants to test themselves more. However, we did not want to create a situation where our participants would have preferred to choose restudy but chose testing anyway. We wanted our participants to want to test themselves while they studied. To do this, we let them decide the difficulty of the test trials.

In Experiment 1, participants could choose pure study trials or pure test trials, but they also had the option to take tests with hints, allowing them to choose a trial type that allowed them to be tested but also get the answer correct. In Experiment 2 the hint options were removed. We predicted that they would try to test themselves as long as they could avoid getting answers wrong. Thus, we predicted that participants would choose test-with-hint trials in Experiment 1, because these trials would allow testing themselves without getting the answers wrong. We predicted that they would choose presentation trials in Experiment 2, when pure tests were the only alternative, because they were likely to get answers wrong if they chose pure test trials. Thus, Experiment 1 and 2 were designed to evaluate self-testing preferences when hints were available (Experiment 1) versus not available (Experiment 2).

Experiment 3 and 4 were designed to examine learning. Our ultimate goal was to improve learning. If hints encourage people to test themselves, that is a good first step. But what if the tests with hints are no more effective than presentations? Then we would not recommend hints as a way to help students learn more efficiently. Hints would only be truly beneficial if they do not destroy the benefits of retrieval. Thus, we examined how much participants learned from pure test trials, hint trials, and pure study trials. In Experiment 3a, we used unrelated word pairs (e.g., idea-seeker) to examine the effects of these hints when the word pairs were not guessable. Experiment 3b was a replication of Experiment 3a with a different population. In Experiment 4, we changed the materials from unrelated word pairs to weakly related word pairs (e.g., whip-punish) to examine the effect of hints when the word pairs were guessable.

Experiment 1

In Experiment 1, participants decided how to study each item during the study phase. On each trial they were shown a cue word (e.g., idea). They controlled how many letters of the target (0, 2, 4, or all 6) were shown. They could choose to see zero letters (e.g., idea-_____), a two-letter condition (e.g., idea-s____r), a four-letter condition (e.g., idea-se__er), or all six letters (e.g., idea-seeker). The first of these trial types is a pure test, the next two are tests with hints, and the last is a pure presentation. We predicted that participants would take advantage of the hints to ensure they would have a good chance of getting the answers correct during the study phase, and in doing so would prefer some sort of test trial over pure presentation. In other words, we offered participants hints so that they would engage in retrieval practice.

Method

Participants

We calculated power by assuming a medium size effect ($\eta^2_p = .06$) based on previous studies that have examined preferences for testing versus restudy. We needed 30 participants to achieve a power = .90, with alpha = .05. We requested more than this to account for participants who did not finish the study as requested.

Fifty-one participants started the experiment. Of these, 10 restarted part-way through and three people did not finish. These participants were excluded from subsequent analyses. Of the remaining 38 participants, one participant indicated that he should be excluded from the analyses because there were major problems during the study. Additionally, we excluded one person who failed to correctly copy the target words during the study phase (mean copy performance was only 1.7%).