

Cognitive Psychology

Assessing the Replicability of Cognitive Psychology During Remote Experiential Learning via Mobile Phone Technology

Benjamin Dyson¹ ^a

¹ Department of Psychology, University of Alberta, Edmonton, AB, Canada

Keywords: Cognitive Psychology, Replication crisis, Open Science, remote learning

<https://doi.org/10.1525/collabra.88320>

Collabra: Psychology

Vol. 9, Issue 1, 2023

A recent global health crisis demanded the wholesale configuration of both teaching and research from in-person to on-line formats. This allowed for an environmental sweep regarding the replicability of some classic and contemporary findings in Cognitive Psychology in the context of an undergraduate course, in which eight portable experimental packages were written for mobile phone. Running across three semesters (average n per study = 585), data consistently produced evidence either for (*Faces, Search, Object, RPS, Rotate*) or against (*Doodle, Trivia*) the original findings, with the exception of one study (*House*) that produced ambiguous findings. The scheme not only allows students exposure to and discussion of the replication crisis within empirical science, but also provides a framework for the future implementation of experiential learning during remote and asynchronous teaching. With continued evaluation made possible via Open Science Framework, a central question is whether on-line data collection violates an essential auxiliary assumption for the replication of in-person data.

Fundamental questions related to the differences between in-person and on-line formats for both higher-education teaching and empirical research were recently placed center stage as a result of a global health crisis. The efficacy of on-line teaching (Ma et al., 2021), the costs and benefits of synchronous versus asynchronous delivery (Fabriz et al., 2021), and, how to implement remote experiential learning (Gallegos et al., 2022) have all become questions of pragmatic necessity. In addition to the wholesale reconfiguration of teaching practices, this also forced the hand of many laboratories collecting in-person human data to switch to on-line data collection. Discussions regarding the relationship between in-person and on-line data quality were also not new, but- like the translation of teaching practices- many researchers were considering the myriad of issues for the first time (Nelson et al., 2021). For example, as a vehicle for on-line data collection, Mechanical Turk (*MTurk*; Buhrmester et al., 2018) has been the focus of scrutiny with respect to potential differences in motivation (Litman et al., 2015) and self-reported well-being (Stone et al., 2019) within its population. At a more pragmatic level, the loss of control regarding the conditions under which participants contributed data removed the high degree of internal validity often championed by human laboratory-based studies (eg, Anderson et al., 1999).

Like many, the potential disruption of moving both teaching and research on-line were compounded in a second-year Cognitive Psychology course at the University of Alberta. Historically, the class collectively attempted to replicate classic and contemporary findings (1971 - 2016) within the literature as an in-person exercise. This curriculum feature has four positive aspects. First, such demonstrations provide opportunities for experiential learning by bringing empirical papers to life via direct activity in the classroom (cf., 'interactive windows'; Dyson, 2008; Huxham, 2005). Second, students selected one of the resultant data sets (and the related target article) to serve as the basis for their written assignment, thereby incorporating the positive feature of student-selected materials into summative assessment (eg., Johnson & Blair, 2003). Third, comparisons between the results expected from the previous literature and those observed within class also facilitated a discussion of the replication crisis within Psychology (after Chopik et al., 2018; Open Science Collaboration, 2015; Pashler & Harris, 2012; Wiggins & Christopherson, 2019). Fourth, the clear disconnect between the rigorous methodological details reported in the previous literature and the more 'ecological' way such studies were run in-class also generated discussion about the design conditions sufficient to claim that the replication was adequate (Feest, 2019).

^a Correspondence should be addressed to: Ben Dyson, Department of Psychology, P-217 Biological Sciences Building, University of Alberta, Edmonton, AB, T6G 2E9, Canada. E-mail : bjdyson@ualberta.ca.

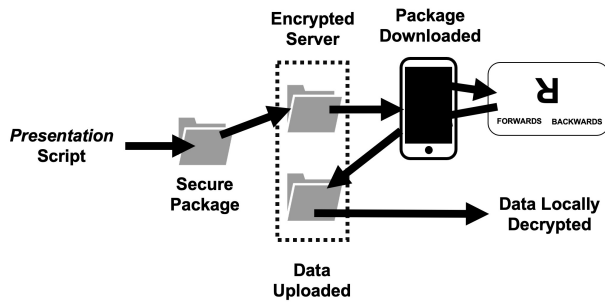


Figure 1. Schematic showing the data flow for delivering a weekly Flex Lab in the context of an undergraduate Cognitive Psychology course using the experimental platform Presentation, its related mobile app Presentation Mobile, and their encrypted servers. Open folder icon by Delapouite under CC BY 3.0 (<https://game-icons.net/>)

To ensure students continued to have experience with empirical studies in a remote learning environment, scripts were written using the experimental software platform *Presentation* (neurobs.com). These scripts were then converted to a secure package format and uploaded to encrypted *Presentation* servers (see [Figure 1](#)) such that they could be further used by the related mobile phone application *Presentation Mobile*. Once students had downloaded the *Presentation Mobile* app. (Android [<https://play.google.com/store/apps/details?id=com.neurobs.presentation&hl=en&pli=1>] or iOS [<https://apps.apple.com/us/app/presentation-mobile/id1163363575>]), they were able to further download each weekly package using search terms and passcodes provided. To use the Winter 2022 semester as an example, individual packages were routinely posted on Monday morning with an e-mail to students that a new study was available. Students could then access the study at any point until Sunday evening; record of participation was noted via students inputting their numeric University of Alberta enrollment ID. The following Monday, the old study was removed and the encrypted data downloaded and locally decrypted. Data were processed the same day and discussion of the study was integrated into a remote interaction session discussing course materials on Tuesday. 9 weekly exercises (known as *Flex Labs*) were run as part of the PSYCH258 course across three semesters: Winter 2021, Fall 2021, and, Winter 2022. All protocols were approved by Research Ethics Board 2 at the University of Alberta (REB2; Pro00103400).

In terms of syllabus design considerations, a number of features were key. First, students had flexible access (Little & Francis, 2005) to the *Flex Labs* in that they could interact with *Presentation Mobile* at any time during the week. Second, in the majority of cases, the *Flex Lab* delivered on Week x was directly related to the course materials to be discussed for Week $x+1$, thereby facilitating distributed learning (Litman & Davachi, 2008). For example, a *Flex Lab* examining the relationship between hemispheric specialization and emotion was completed the week before course materials on *Cognitive Neuroscience*. Third, participation marks (1% of course grade) were allocated to each

Flex Lab, with a bonus 1% for engagement with all 9 *Flex Labs*, thereby providing external motivation for course participation (Gallo & Rinaldo, 2011). Targets for replication were largely driven by two pragmatic concerns: a) the presence of existing materials used in previous iterations of in-person versions of the Cognitive Psychology course, and, b) the pressures in converting these materials to remote experimental packages such that they were ready on time for our first fully on-line semester of Fall 2021 (used as a pilot semester) prior to more formal data collection in Winter 2021.

Each of the resultant 8 *Flex Labs* data sets across three semesters of on-line teaching are now summarized with respect to: their relationship with the course materials, the methodology used, the average effect sizes generated either in support or in refutation of the original observation across three semesters, and, a comparison with the original effect size (when available). [Table 1](#) provides a summary of effects. Fixed effects models were used since it was assumed the true effect size should be the same across all three points of data collection (Winter 21, Fall 21, Winter 22). Random effect models are preferred when there is heterogeneity across data collection (Barili et al., 2018). Experimental packages and summary data freely available via the Open Science Framework (<https://osf.io/kv5qp/>). The *Flex Lab* data illustrate a number of points related to the empirical health of Cognitive Psychology in the context of the replication crisis, tackled in the *General Discussion*.

Studies

Experiment 1 (Faces)

Within the context of *Cognitive Neuroscience*, Levy, Heller, Banich & Burton (1983) discuss the potential lateralization of emotional processing, using a chimeric face procedure. Chimeric faces are created by vertically cutting and pasting two versions of the same face: one with and one without emotion. As such, two copies of each face exist, one where the left side is neutral and right side emotional, and, where the right side is emotional and the left side neutral. Levy et al (1983; see also Rueckert & Naybar, 2008) inferred a preference for emotional processing in the right hemisphere by revealing a left-sided bias when participants were asked to choose between two copies of chimeric faces.

On each of 36 trials, a lateralized pair of chimeric faces taken from Levy et al. (1983) were presented. Each faces juxtaposed a neutral side and happy (smiling) side. On 18 trials, when emotion appeared on the left-side of the chimeric face, this face was presented on the left. On 18 trials, when emotion appeared on the left-side of the chimeric face, this face was presented on the right. Trial order was individually randomized for each participant. Responding was self-paced, with participants pressing either the left or right side of their device to “select the face that you think looks the happiest.” After 36 trials, participants were indicated whether they were left- or right-handed. The degree of bias was calculated from the formula for Laterality Quotient (LQ; Rueckert & Naybar, 2008), which was $[(\text{right responses} - \text{left responses}) / 36]$. As shown in [Figure 2a](#), across all

Table 1. Summary of replication attempts for eight studies providing supporting (Yes), refuting (No) or ambiguous (?) evidence across three semesters (Winter 21, Fall 21, Winter 22)

		Winter 21	Fall 21	Winter 22
Faces	(Levy et al., 1983) There is a left-side bias for chimeric faces when choosing the face that looks happier	Yes	Yes	Yes
Search	(Wolfe, 2001) Reaction times are faster for unknown targets defined by the presence (not absence) of a feature	Yes	Yes	Yes
Doodle	(Andrade, 2010) Doodling increases intentional memory related to an pre-recorded audio message	No	No	No
House	(Pichert & Anderson, 1976) Switching perspective provides new information related to a previous passage	Yes	?	?
Object	(Brady et al., 2008) There is better than chance recognition of old objects presented in the context of state pairs	Yes	Yes	Yes
Trivia	(Pashler et al., 2005) Performance following 'correct' / 'incorrect' feedback is worse than providing no feedback	No	No	No
RPS	(Dyson et al., 2016) The percentage of shift behaviour is larger following losses relative to wins	Yes	Yes	Yes
Rotate	(Cooper & Shepard, 1973) Letter categorization takes longer as the letter is rotated away from its canonical position	Yes	Yes	Yes

three semesters, a small effect (Cohen, 1988) was consistently observed in support of this hypothesis, resulting in an average effect size estimate of 0.262 [CI +/- 95%: 0.177 – 0.347; $n = 551$]. This is relative to the large effect originally reported by Levy et al. (1983) of $d = 0.689$ [CI +/- 95%: 0.480 – 0.894; $n = 111$].

Experiment 2 (Search)

For the topic of *Perception*, the visual search paradigm and feature integration theory (FIT; Treisman & Gelade, 1983) were introduced. As summarized by Wolfe (2001), asymmetries exist during visual search performance, in that it is easier to detect unknown targets defined by the presence rather than absence of a feature. For example, finding a Q amongst a series of Os is easier than finding an O amongst Qs, since the Q contains an additional diagonal stroke.

On each of 24 trials, participants were presented with a visual search display, containing 20 spatially diffuse elements. 19 of those elements were identical (distractors) and 1 was unique (target). The stimuli used included O vs. Q, forwards vs. backwards letters, O vs. C, rotated vs. non-rotated letters, and, diagonal vs. vertical lines. For 12 *feature present* trials, one element was assigned target and the other element was assigned distractor (e.g., 1 Q, 19 Os). For the other 12 *feature absent* trials, the assignment of element to target and distractor was switched (e.g., 1 O, 19 Qs). Trial order was individually randomized for each participant. Responding was self-paced, with participants asked to “as fast as possible, press the odd one out.” A robust, large effect size was consistently observed in support of the faster de-

tection of a single target defined by the presence rather than absence of a feature (average effect size estimate = 0.887 [CI +/- 95%: 0.799 – 0.976]; $n = 668$; see [Figure 2b](#)).

Experiment 3 (Doodle)

Andrade (2010) reports a novel finding of dual-task benefit rather than dual-task decrement, in the context of *Attention*. Specifically, participants were asked to monitor a telephone message and to note the names of individuals who would either definitely or probably be coming to a hypothetical party. In contrast to a *single task* group, participants in another group (*dual task* group) were also asked to ‘doodle’ (fill in shapes) as they were listening to the message. Not only were the doodle group observed to have better intentional memory of the party-goers (Andrade, 2010, $n = 40$; *Hedges g* = 0.826 [CI +/- 95%: 0.094, 1.377]), but in a surprise memory test, the doodle group also performed better at incidentally remembering locations that were mentioned during the message (ibid; *Hedges g* = 0.416 [CI +/- 95%: 0.205, 1.049]).

A variant of the original Andrade (2010) message was recorded, specifically replacing UK city locations with major Canadian cities. Participants were allocated to either the non-doodle (*single-task*) or doodle (*dual-task*) group based on the first letter of their last name (A-M, or, N-Z, respectively). Participants in the dual-task group were further instructed to download a version of the Doodle sheet. Both groups were then directed to https://www.youtube.com/watch?v=YUbZ_Owp4c. The video contained instructions based on Andrade (2010), and an audio reading of the modified phone message (approximately 2m30s) during which

time, participants were encouraged to note who would likely be coming to the party. An incidental memory prompt then appeared for approximately 1m, asking participants to write down any city names they could remember. All individuals were then directed to Presentation Mobile. This script prompted them to record their group allocation, followed by the presentation of 8 names and 8 places. For each of the 16 trials, they were asked to indicate whether they had remembered each specific item. Trial order within each category was individually randomized and responding was self-paced. Our data reliably produced data contrary to the observation that intentional memory was facilitated by concurrent doodling, as indexed by a negative average effect size = -0.340 [CI +/- 95%: -0.497 – -0.182] ($n = 653$; see [Figure 2c](#)).

Experiment 4 (House)

The role of schema in the topic of *Everyday Memory* was explored using the study by Pichert & Anderson (1976). Here, participants were provided with a perspective from which to imagine a scenario regarding the details of a house: either imagine the passage as a *buyer* or *burglar* of the house. When tested on buyer-related and burglar-related items contained within the passage, the influence of perspective on memory was demonstrated in that perspective-consistent information (e.g., *buyer* – leaking roof) was better remembered than perspective-inconsistent information (e.g., *burglar* – leaking roof). Furthermore, when participants were asked to switch rather than maintain perspectives, this ‘unlocked’ new information about the passage unavailable at the time of initial retrieval.

Participants were assigned to 1 of 4 groups on the basis of the first letter of their last name (A-D, *Buyer-Buyer*; E-K, *Buyer-Burglar*; L-P *Burglar-Buyer*; Q-Z, *Burglar-Burglar*). Groups were then directed to watch a tailored YouTube video, where they were instructed to listen to an audio version of the Pichert & Anderson (1976) narrative (approximately 2 minutes) from a particular perspective (*Buyer* or *Burglar*), eg, Group A: <https://www.youtube.com/watch?v=pASKfnJ00O4>. Participants then had two minutes to “write down as much of the exact story as you can.” This was followed by another two minute tailored prompt that either asked participants to continue to recall the story from the same perspective (*Buyer – Buyer*, *Burglar – Burglar*), or, to switch into a different perspective (*Buyer – Burglar*, *Burglar – Buyer*). All individuals were then directed to Presentation Mobile and prompted to record their group allocation. This was followed by the presentation of 12 *Buyer* relevant items and 12 *Burglar* relevant items in a random order, with participants responding as to whether they had remembered them. Trial order within each category was individually randomized and responding was self-paced. To finish the study, participants were invited to completed both the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova et al., 2006). In testing the hypothesis that switching perspectives increased the recall of initially irrelevant information, the data across the three semesters of PSYCH258 were highly variable (see [Figure 2d](#); $n = 619$),

resulting in an ambiguous average effect size = -0.016 [CI +/- 95%: -0.188 – 0.155]. The reasons for this variability will be discussed later.

Experiment 5 (Object)

A perennial question regarding *Long-Term Memory* is its limit: exactly what level of detail is retained within memory, and, for how long? Brady et al. (2008) provided provocative data suggesting that high fidelity details regarding visual objects are retained within long-term memory, even when those objects are only viewed once. Specifically, 2896 objects were presented across a 5-6 hour period, in which participants were required to count the number of object repetitions that occurred. Following that lengthy period of encoding, participants were shown a further 300 pairs of objects, with one object in the pair being previously seen. Brady et al. (2008) were able to assess the level of detail retained within long-term memory by manipulating the degree of similarity between old and new objects. In novel pairs, the new object was completely new (e.g., a chocolate cake versus a toy train). In exemplar pairs, the new object was a different version of the object (e.g., two chocolate cakes). In state pairs, the new object was the old object in a different configuration (e.g., the same chocolate cake but with a slice taken out). Memory performance for novel, exemplar and state pairs was 93%, 88% and 87%, respectively.

Data collection took part across two separate weeks of the course. For the first week, participants were shown a sample of 144 images from Brady et al. (2008) at a rate of 1 every 3 seconds. 120 images were shown only once, whereas 12 images were shown twice ($2 \times 12 = 24$). During this initial phase of the study, participants were explicitly asked to count the number of image repetition they think they saw. To end their contribution, they selected a range that contained the number of repetition they thought they saw (0-2, 3-5, 6-8, 9-11, 12-14, 16-17). These repeated objects played no further role in the study.

For the second week of data collection, participants had to indicate whether they participated in the first week. Regardless of their response, participants were then shown 120 pairs of images. In each pair, there was an image they had seen during the first week along with a new image. This new image was a novel object, a different exemplar of the same object category, or, the same object in a different state. There were 40 trials of each of the 3 kind of pairs. Responding was self-paced and image pairs were presented in individual random order. To move on to the next trial, participants selected the side of the screen containing the old object (and were encouraged to guess). Although a straightforward replication of Brady et al. (2008) was impossible given the constraints of teaching, re-testing 120 objects across a two week interval revealed a small reliable effect in favour of better than chance recognition of old objects in the context of state pairs: average effect size estimate = 0.314 [CI +/- 95%: 0.227 – 0.400; $n = 543$; see [Figure 2e](#)].

Experiment 6 (Trivia)

Key practicalities arising from the course discussion of *Semantic Memory* are the ways in which memory can be better remembered. Specifically in the context of education, it is a pertinent question what kind of feedback enables memory improvement. Pashler, Cepeda, Wixted & Rohrer (2005) examined three different classes of feedback over a variety of retention periods. In response to learning a novel Lugandan dialect, they compared no feedback, with the feedback of being ‘correct’ or ‘incorrect’, with the provision of the correct answer. They found that performance following ‘correct’ / ‘incorrect’ messaging was actually *worse* than no feedback, but that providing the correct answer had a benefit over no feedback. Intriguingly, the authors report that the “...basic pattern of results described above was confirmed in another, quite similar, online experiment that we carried out teaching subjects obscure facts rather than foreign language vocabulary, and using a within-subject instead of a between-subjects manipulation of the type of feedback.” (Pashler et al., 2005, p. 6). Inspired by this comment, I asked whether providing the feedback labels of ‘correct’ and ‘incorrect’ would have similarly detrimental effects, relative to a no feedback condition in the context of updating trivia knowledge.

60 trivia questions were identified from a variety of online source encompassing topics including Science & Nature, Religion & Mythology, Food & Drink, Art & Literature. Participants were informed that they were invited to answer 120 trivia questions: 60 at Stage 1 and 60 at Stage 2. For each Stage 1 trial, participants were shown a question with two responses: one on the left and one on the right. After pressing which response they thought was correct, they were provided with no feedback, the feedback of ‘correct’ or ‘incorrect’, or, provided with the correct answer. 20 questions were allocated to each of these 3 feedback categories, with trial order individually randomized. At the beginning of Stage 2, participants were informed that they were invited to answer the same 60 questions again, and that in all cases feedback would be delivered in the form of ‘correct’ or ‘incorrect’. We reliably produced a large effect size contrary to this claim (-0.951 [CI +/- 95%: -1.072 - -0.831]; $n = 572$; see [Figure 2f](#)).

Experiment 7 (RPS)

In our discussion of *Decision Making*, the class was introduced to how human decisions can be sub-optimal. To demonstrate this concept, the three-response zero-sum game of Rock, Paper, Scissors was used. Here, the only way to guarantee not being beaten in the game is to play according to mixed-strategy. This is where individual items are played with equal probability, and there is no association between the previous outcome and current response (Dyson et al., 2016). However, operant conditioning principles such as *lose-shift* continue to influence game performance in sub-optimal ways. Specifically, Dyson et al. (2016) noted that the probability of changing responses is more likely after losing relative to winning a RPS round (ie,

lose-shift > *win-shift*; (Hedges $g = 0.788$ [CI +/- 95%: 0.379, 1.228]).

90 trials of Rock, Paper, Scissors were delivered over the Presentation Mobile platform. Following instructions, participants tapped a left, middle or right location to indicate a Rock, Paper or Scissors response. Following a 500 ms pause, both computer (left) and participant (right) response were shown for 1000 ms (using stimuli adapted from Forder & Dyson, 2016). Following another 500 ms pause, trial outcome was reported as ‘WIN (+1)’, ‘LOSE (-1)’ or ‘DRAW (0)’ for 1000 ms. Scores were updated during a final 500 ms interval, and another response was requested. The computerized opponent played 30 Rock, 30 Paper, 30 Scissors responses in an individually randomized order. A small effect size consistent with larger *lose-shift* relative to *win-shift* proportions was also observed across the three semesters of PSYCH258 (0.295 [CI +/- 95%: 0.209 - 0.382]; $n = 560$; see [Figure 2g](#)).

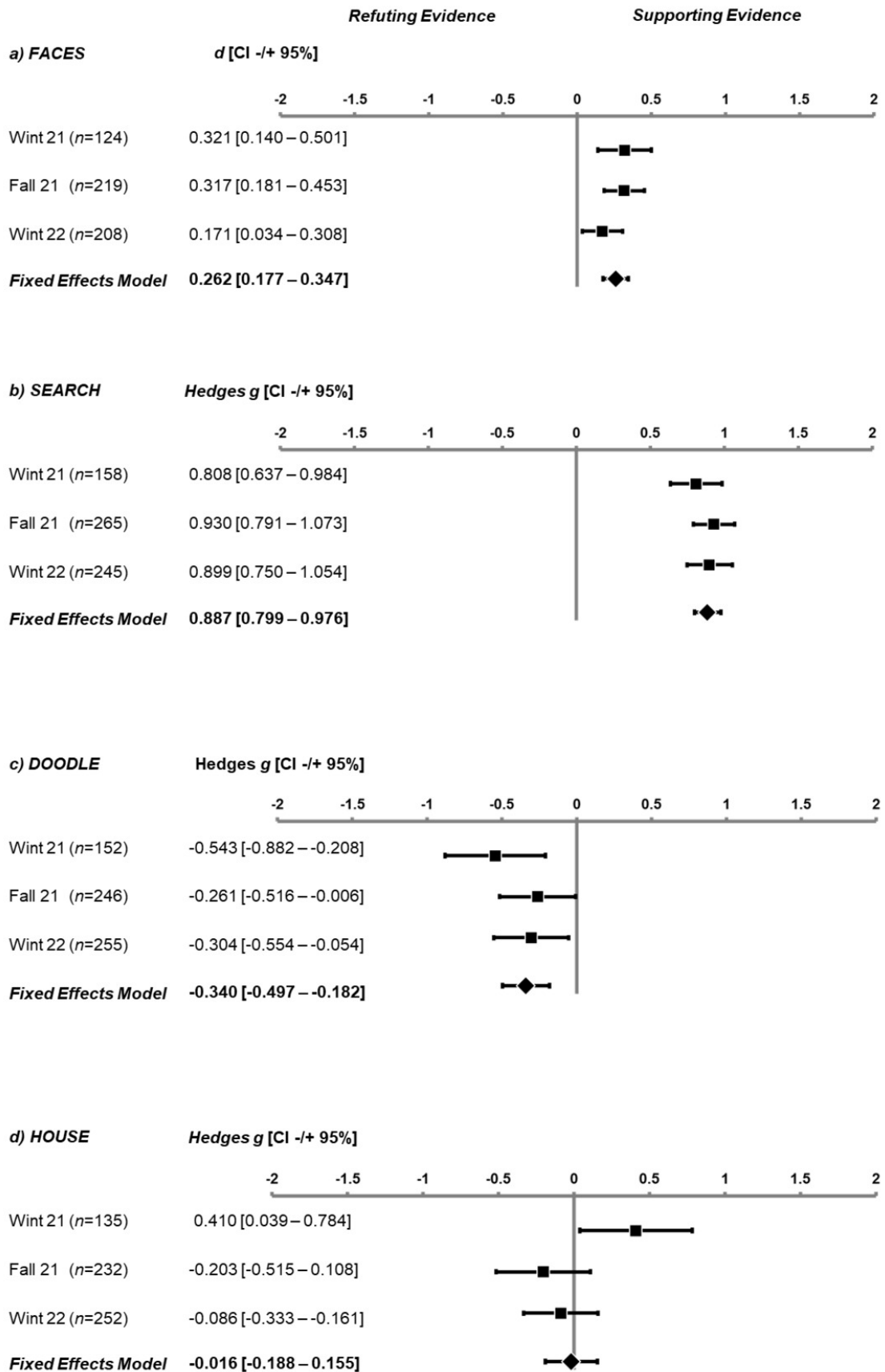
Experiment 8 (Rotate)

A final, classic Cognitive Psychology finding relates to *Visual Imagery* and behavioural data consistent with the notion of mental rotation. From Cooper & Shepard (1973), participants were asked to categorize shapes with varying degrees of rotation, as either forwards-facing or backwards-facing letter Rs. Of principle concern was whether the degree of effort involved in physical rotation was akin to the degree of effort involved in mental rotation. If this was the case, then the further a shape was rotated away from its canonical position (i.e. 0 degrees), the longer it should take to rotate back in order to confirm letter categorization.

12 images were shown via Presentation Mobile: a forward-facing R or backwards-facing R (categorization x 2), rotated by 0, 60, 120, 180, 240, 300 degrees (rotation x 6). Each image was shown 15 times in a random order, resulting in 180 trials in total. Participants were asked to be both fast and accurate, pressing the left side of their screen if the shape was a forwards-facing R, and, the right side of their screen if the shape was a backwards-facing R. Feedback was not provided, and a 500 ms pause was introduced between each stimulus presentation. Only correct RTs as a function of rotation were considered in analysis. Reliably large effect sizes were shown across Winter 2021, Fall 2021 and Winter 2022 samples (average effect size = 1.039 [CI +/- 95%: 1.009 - 1.070]; $n = 518$; see [Figure 2h](#)).

Discussion

The recent requirement to move both teaching and research from in-person to on-line presented an opportunity to conduct an environmental sweep on the empirical health of Cognitive Psychology. In the context of an undergraduate course, 8 empirical studies were selected- each representing a classic or contemporary observation directly linked to course materials- and an on-line version of each study (*Flex Lab*) was implemented via *Presentation Mobile*. Running across three semesters, the average sample size for each study was 585- well exceeding the recommendation that attempted replications should have a sample size



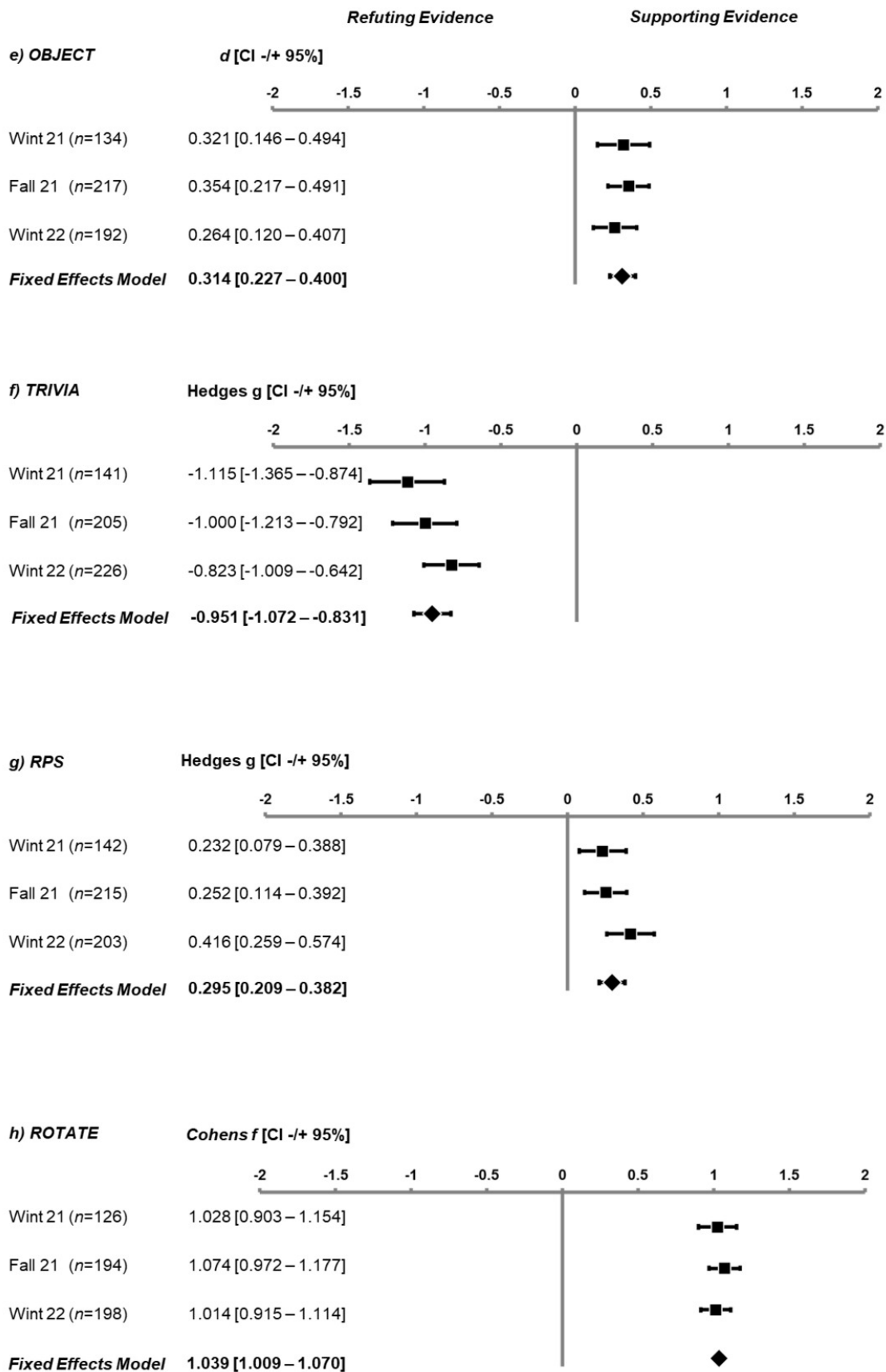


Figure 2. Individual effect size estimates for all 8 studies across three semesters Winter 2021, Fall 2021 and Winter 2022, in addition to an estimate of average effect size based on a fixed effects model (Turner & Bernard, 2006): a) Faces ($n = 551$), b) Search ($n = 668$), c) Doodle ($n = 653$), d) House ($n = 619$), e) Object ($n = 543$), f) Trivia ($n = 572$), g) RPS ($n = 560$), and, h) Rotate ($n = 518$),

2.5 times the original (Simonsohn, 2013), when such information was available.

There were two *Flex Labs* (*Faces*, *RPS*) where an estimate of the original effect size was available and the current data provided support for the original hypothesis. The current effect size was at least half the size of the original effect size (*Faces*, Levy et al. (1983) $d = 0.689$ versus current $d = 0.262$; *RPS*, Dyson et al. (2016) $g = 0.788$ versus current $g = 0.295$). Such observations are consistent with the likelihood that published effect sizes are an overestimate of true effect size, since- historically- studies finding non-significant results are not published to the same degree (Nuijten et al., 2015; Schäfer & Schwarz, 2019; see also Serragarcia & Gneezy, 2021). Furthermore, two additional *Flex Labs* (*Search*, *Rotate*) may seem relatively uncontroversial in terms of providing both robust and large effect sizes in support of speeded reaction time for the detection of unknown targets defined by feature presence rather than absence (*Search*; after Wolfe, 2001), and, slowed reaction time as the degree of mental rotation increases (*Rotate*; after Shepard & Metzler, 1971). The observation of such robust and large effect sizes might lead us to place some confidence in the empirical health of certain key (albeit selective) findings within Cognitive Psychology (Open Science Collaboration, 2015; see their Table 1). A cautionary note is warranted however in terms of the absence of accuracy data for the *Search* lab. Similarly, in the case of *Rotate*, the extent to which speed is sacrificed for accuracy during mental rotation remains an open question (eg, Liesefeld et al., 2015).

There were also two cases (*Doodle*, *Trivia*) where current data reliably provided evidence refuting the original hypothesis. For the *Doodle Flex Lab*, we consistently found evidence for a dual-task decrement rather than dual-task benefit with respect to the auditory memory of a telephone message and the visuo-motor act of doodling (after Andrade, 2010). This discrepancy highlights the difficulties in reconciling attempts at replication that yield contrary observations in terms of the inclusion or exclusion of auxiliary assumptions. Hudson (2021) provides the example where a researcher erroneously assumes laboratory temperature is an irrelevant aspect of a replication design, and then fails to find evidence in favour of the original hypothesis as a result of setting laboratory ambient temperature at -25°C (see also Feest, 2019, for a similar example). Undoubtedly, there were similar- albeit less extreme- assumptions made in re-implementing Andrade (2010). First, the original study (conducted in the UK) contained both putatively Caucasian names (eg, Claire, William) and UK towns (eg, Colchester, Peterborough) as the target of monitored and incidental recall, respectively. Since the current version was completed in Canada, it was considered important to change UK towns to Canadian cities (eg., Montreal, Vancouver). Second, and because of the first, the audio message provided in the original Appendix was re-recorded. Third, participants in the original Doodle study were asked to complete this study directly after completing another, as a way to increase boredom “by testing people who were already thinking about going home.” (Andrade, 2010, p. 101).

This is in direct contrast to the current study, where it could be argued individuals were actively seeking participation. Thus, if boredom is a key mechanism driving the effect, I have failed to adequately represent the original study. Fourth, in the original study participants were paid for participation (as opposed to given course grade). In sum, a failure to replicate may always be attributed to changes in auxiliary assumptions. A central question for all future research is whether in-person rather than on-line data collection is an essential auxiliary assumption for replication. The interested reader is also directed to Murre’s (2021) recent failure to similarly replicate Godden & Baddeley’s (1975) ‘classic’ finding of context-dependent memory between being on land and underwater. To Murre’s (2021) credit, a number of auxiliary assumptions (of unknown importance) are noted in the attempted replication: the use of an indoor pool versus open water, the amateur or professional status of the divers, testing divers in a single day relative to across 4 days, to name but three.

In the case of the *Trivia Flex Lab*, we again found evidence contrary to the position of Pashler et al. (2005) that the delivery of ‘correct / incorrect’ feedback was detrimental to the retention of trivia information relative to the provision of no feedback. Here, however, the degree to which the inclusion or exclusion of auxiliary assumptions may determine the direction of the effect cannot be properly assessed due to an absence of detail in the original study itself. The footnote reported in the Results section above (Pashler et al., 2005, p. 6) is the extent of detail related to the original study. Hence, exact (or, direct) replications are never possible for studies for which original details cannot be retrieved. As such, replications are immediately restricted to the category of experimental or conceptual replications (see Hudson, 2021). As such, we offer our *Trivia* package (and all other packages) via the Open Science Framework to enable a direct replication of this work, itself reflecting a conceptual replication of the original Pashler et al. (2005) study.

A pragmatic note on some difficulties associated with implementing on-line versions of paradigms. It should be noted that for the *Object Flex Lab*, neither a direct nor a conceptual replication of the Brady et al (2008) work was possible due to the significant time demands of the original study (approximately 6 hours). Hence, we are left with a different demonstration that- over a longer time period with a smaller sample of objects- individuals are able to remember specific object states at a level greater than chance. This is opposed to (but does not necessarily contradict) the original conclusion of the Brady et al (2008) study, where long-term memory for novel, exemplar and state objects was close to ceiling. Finally, in the case of the Winter 2021 and Fall 2021 *House* lab data, it was apparent that the script failed to accurately record group allocation, due to errors in touch screen allocation. Thus, the script for the *House Flex Lab* was re-written for the Winter 2022, rendering the comparison between semesters ambiguous.

In summary, this project attempted to meet the challenges of the configuration of both Cognitive Psychology teaching and research to on-line formats. There are nu-

merous ways in which Open Science and reproducibility or replicability can be integrated into higher education, yet all share the ultimate goals of positively affecting students' engagement, attitudes toward science and scientific literacy (see Pownall et al., 2023, for a review). For example, the Collaborative Replications and Education Project (CREP) takes a student-led rather than instructor-led approach to the preparation of attempted replication materials (see Wagge et al., 2019, for an example). The current solution of how to provide experiential, replication crisis learning during remote and asynchronous delivery was provided by the development of portable experimental packages written for mobile phone via the *Presentation Mobile* app. These *Flex Labs* represented attempted replications of classic and contemporary findings within the literature, which also explicitly connected with course material. The resultant data not only further allowed students to question the degree to which replication is possible within empirical science but also potentially contribute to the Cognitive Psychology literature in highlighting both robust (eg., *Rotation*; cf. Shepard & Metzler, 1971) and potentially ambiguous (*Trivia*; cf. Pashler et al., 2005) previous findings. The *Flex Lab* program provides a framework for supporting future replication attempts utilizing large on-line samples both in and beyond Cognitive Psychology, with the precise operationalization of the study preserved by the open access availability of study packages (<https://osf.io/kv5qp/>).

Contributions

B.J.D. designed the studies, analysed the data and wrote the manuscript.

Acknowledgements

The authors would like to thank Dr Peggy St Jacques (University of Alberta, Canada), Dr Julia Spaniol (Toronto Metropolitan University, Canada) and Dr Graham Hole (Sussex University, UK) for providing some the materials related to the studies, and, Ver-Se Denga for providing the trivia questions. The author would also like to acknowledge the Teaching Assistants for PSYCH258: Yafei Qi & Eamin Zahan Heanoy (Fall 2020), Michelle Tomazck & Arturo Perez (Winter 2021), Sijie Ling and Eamin Zahan Heanoy (Fall 2021), and, Eunchan Na and Yajing Zhang (Winter 2022). Part of the data were presented at the 2021 and 2022 Festival of Teaching and Learning, University of Alberta, Canada.

Funding Information

There was no funding.

Competing Interests

The author declares no competing interests.

Supplemental Material

Materials for all experiments are available at <https://osf.io/kv5qp>.

Data Accessibility Statement

Summary data for all experiments are available at <https://osf.io/kv5qp>.

Submitted: March 01, 2023 PDT, Accepted: August 07, 2023 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, 8(1), 3–9. <https://doi.org/10.1111/1467-8721.00002>
- Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology*, 24(1), 100–106. <https://doi.org/10.1002/acp.1561>
- Barili, F., Parolari, A., Kappetein, P. A., & Freemantle, N. (2018). Statistical primer: Heterogeneity, random- or fixed-effects model analyses? *Interactive Cardiovascular and Thoracic Surgery*, 27(3), 317–321. <https://doi.org/10.1093/icvts/ivy163>
- Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, 20(2), 239–263. <https://doi.org/10.1002/acp.1182>
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, 105(38), 14325–14329. <https://doi.org/10.1073/pnas.0803390105>
- Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, 13(2), 149–154. <https://doi.org/10.1177/1745691617706516>
- Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, 45(2), 158–163. <https://doi.org/10.1177/0098628318762900>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase, *Visual information processing* (pp. 75–176). Academic. <https://doi.org/10.1016/b978-0-12-170150-5.50009-3>
- Dyson, B. J. (2008). Assessing small-scale interventions in large-scale teaching: A general methodology and preliminary data. *Active Learning in Higher Education*, 9(3), 265–282. <https://doi.org/10.1177/1469787408095856>
- Dyson, B. J., Wilbiks, J. M. P., Sandhu, R., Papanicolaou, G., & Lintag, J. (2016). Negative outcomes evoke cyclic irrational decisions in Rock, Paper, Scissors. *Scientific Reports*, 6(1), 20479. <https://doi.org/10.1038/srep20479>
- Fabriz, S., Mendzheritskaya, J., & Stehle, S. (2021). Impact of synchronous and asynchronous settings of online teaching and learning in higher education on students' learning experience during COVID-19. *Frontiers in Psychology*, 12, 733554. <https://doi.org/10.3389/fpsyg.2021.733554>
- Feest, U. (2019). Why replication is overrated. *Philosophy of Science*, 86(5), 895–905. <https://doi.org/10.1086/705451>
- Gallegos, P. J., Hoffmaster, B. S., Howard, M. L., Lancaster, J. W., Pluckrose, D., Smith, B. A., Tallian, K., Van Matre, E. T., & Scott, J. D. (2022). Remote experiential education: A silver lining from the COVID-19 pandemic. *Journal of the American College of Clinical Pharmacy*, 5(1), 107–110. <https://doi.org/10.1002/jac5.1572>
- Gallo, M., & Rinaldo, V. (2011). Intrinsic versus extrinsic motivation: A study of undergraduate student motivation in science. *Teaching and Learning*, 6(1), 95–106. <https://doi.org/10.26522/tl.v6i1.379>
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, 66(3), 325–331. <https://doi.org/10.1111/j.2044-8295.1975.tb01468.x>
- Hudson, R. (2021). Explicating exact versus conceptual replication. *Erkenntnis*, 88(6), 2493–2514. <https://doi.org/10.1007/s10670-021-00464-z>
- Huxham, M. (2005). Learning in lectures: Do 'interactive windows' help? *Active Learning in Higher Education*, 6(1), 17–31. <https://doi.org/10.1177/1469787405049943>
- Johnson, D., & Blair, A. (2003). The importance and use of student self-selected literature to reading engagement in an elementary reading curriculum. *Reading Horizons: A Journal of Literacy and Language Arts*, 43, 3.
- Levy, J., Heller, W., Banich, M. T., & Burton, L. A. (1983). Asymmetry of perception in free viewing of chimeric faces. *Brain and Cognition*, 2(4), 404–419. [https://doi.org/10.1016/0278-2626\(83\)90021-0](https://doi.org/10.1016/0278-2626(83)90021-0)
- Liesefeld, H. R., Fu, X., & Zimmer, H. D. (2015). Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 1140–1151.
- Litman, L., & Davachi, L. (2008). Distributed learning enhances relational memory consolidation. *Learning & Memory*, 15(9), 711–716. <https://doi.org/10.1101/11132008>
- Litman, L., Robinson, J., & Rosenzweig, C. (2015). The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. *Behavior Research Methods*, 47(2), 519–528. <https://doi.org/10.3758/s13428-014-0483-x>
- Little, E., & Francis, A. (2005). Teaching introductory psychology through flexible delivery: a case study. *Psychology Learning and Teaching*, 5, 47–41.
- Ma, K., Chutiyami, M., Zhang, Y., & Nicoll, S. (2021). Online teaching self-efficacy during COVID-19: Changes, its associated factors and moderators. *Education and Information Technologies*, 26(6), 6675–6697. <https://doi.org/10.1007/s10639-021-10486-3>

- Marks, D. F. (1973). Visual imagery differences in the recall of picture. *British Journal of Psychology*, *64*(1), 17–24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>
- Murre, J. M. J. (2021). The Godden and Baddeley (1975) experiment on context-dependent memory on land and underwater: a replication. *Royal Society Open Science*, *8*, 200724.
- Nelson, P. M., Scheiber, F., Laughlin, H. M., & Demir-Lira, ö E. (2021). Comparing face-to-face and online data collection methods in preterm and full-term children: An exploratory study. *Frontiers in Psychology*, *28*, 733192.
- Nuijten, M. B., van Assen, M. A. L. M., Veldkamp, C. L. S., & Wicherts, J. M. (2015). The replication paradox: Combining studies can decrease accuracy of effect size estimates. *Review of General Psychology*, *19*(2), 172–182. <https://doi.org/10.1037/gpr0000034>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *346*, aac4716. <https://doi.org/10.1126/science.aac4716>
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. <https://doi.org/10.1037/0278-7393.31.1.3>
- Pashler, H., & Harris, C. R. (2012). Is the replication crisis overblown? *Perspectives on Psychological Science*, *7*(6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Pichert, J. W., & Anderson, R. C. (1976). Taking different perspectives on a story. *Journal of Educational Psychology*, *68*, 309–315.
- Pownall, M., Azevedo, F., König, L. M., Slack, H. R., Evans, T. R., Flack, Z., Grinschgl, S., Elsherif, M. M., Gilligan-Lee, K. A., de Oliveira, C. M. F., GJoneska, B., Kalandadze, T., Button, K., Ashcroft-Jones, S., Terry, J., Albayrak-Aydemir, N., Déchtérenko, F., Alzahawi, S., Baker, B. J., ... Sadhwani, S. (2023). Teaching open and reproducible scholarship: a critical review of the evidence base for current pedagogical methods and their outcomes. *Royal Society Open Science*, *10*(5), 221255. <https://doi.org/10.1098/rsos.221255>
- Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition*, *67*(2), 162–167. <https://doi.org/10.1016/j.bandc.2008.01.002>
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, *10*, 813. <https://doi.org/10.3389/fpsyg.2019.00813>
- Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, *7*(21), eabd1705. <https://doi.org/10.1126/sciadv.abd1705>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. <https://doi.org/10.1126/science.171.3972.701>
- Simonsohn, U. (2013). Small telescopes: Detectability and the evaluation of replication results. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2259879>
- Stone, A. A., Walentynowicz, M., Schneider, S., Junghaenel, D. U., & Wen, C. K. (2019). MTurk participants have substantially lower evaluative subjective well-being than other survey participants. *Computers in Human Behavior*, *94*, 1–8. <https://doi.org/10.1016/j.chb.2018.12.042>
- Treisman, A., & Gelade, G. (1983). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. [https://doi.org/10.1016/0010-0285\(83\)90005-5](https://doi.org/10.1016/0010-0285(83)90005-5)
- Turner, H. M., III, & Bernard, R. M. (2006). Calculating and synthesizing effect sizes. *Contemporary Issues in Communication Science and Disorders*, *33*(Spring), 42–55. <https://doi.org/10.1044/cicsd.33.s.42>
- Wagge, J. R., Baciú, C., Banas, K., Nadler, J. T., Schwarz, S., Weisberg, Y., IJzerman, H., Legate, N., & Grahe, J. (2019). A demonstration of the Collaborative Replication and Education Project: Replication attempts of the red-romance effect. *Collabra: Psychology*, *5*(1), 5. <https://doi.org/10.1525/collabra.177>
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. <https://doi.org/10.1037/teo0000137>
- Wolfe, J. M. (2001). Asymmetries in visual search: An Introduction. *Perception & Psychophysics*, *63*(3), 381–389. <https://doi.org/10.3758/bf03194406>

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/88320-assessing-the-replicability-of-cognitive-psychology-during-remote-experiential-learning-via-mobile-phone-technology/attachment/182096.docx?auth_token=L3R8qIOiPynDjSlPuEe1

Response Letter

Download: https://collabra.scholasticahq.com/article/88320-assessing-the-replicability-of-cognitive-psychology-during-remote-experiential-learning-via-mobile-phone-technology/attachment/182097.pdf?auth_token=L3R8qIOiPynDjSlPuEe1
