Meta-analytic findings of the self-controlled motor learning literature: Underpowered, biased, and lacking evidential value

Brad McKay[1,2], Zachary D. Yantha[1], Julia Hussien[1], Michael J. Carter[2], and & Diane M. Ste-Marie[1]

[1] School of Human Kinetics

University of Ottawa

[2] Department of Kinesiology

McMaster University

Author Note

Brad McKay https://orcid.org/0000-0002-7408-2323

Julia Hussien https://orcid.org/0000-0001-7434-228X

Michael J. Carter https://orcid.org/0000-0002-0675-4271

Zachary D. Yantha https://orcid.org/0000-0003-1851-7609

Diane M. Ste-Marie https://orcid.org/0000-0002-4574-9539

Correspondence concerning this article should be addressed to Brad McKay, 1280 Main Street West, Ivor Wynne Centre Room AB-131 A1, McMaster University, Hamilton ON Canada, L8S 4K1. E-mail: bradmckay8@gmail.com

Abstract

The self-controlled motor learning literature consists of experiments that compare a group of learners who are provided with a choice over an aspect of their practice environment to a group who are yoked to those choices. A qualitative review of the literature suggests an unambiguous benefit from self-controlled practice. A meta-analysis was conducted on the effects of self-controlled practice on retention test performance measures with a focus on assessing and potentially correcting for selection bias in the literature, such as publication bias and p-hacking. First, a naïve random effects model was fit to the data and a moderate benefit of self-controlled practice, $g = .44\,(k = 52, N = 3134, 95\%\,CI\,[.31,\ .56])$, was found. Second, publication status was added to the model as a potential moderator, revealing a significant difference between published and unpublished findings, with only the former reporting a benefit of self-controlled practice. Third, to investigate and adjust for the impact of selectively reporting statistically significant results, a weight-function model was fit to the data with a one-tailed p-value cutpoint of .025. The weight-function model revealed substantial selection bias and estimated the true average effect of self-controlled practice as $g = .107\,(95\%\,CI\,[.047,\ .18])$. P-curve analyses were conducted on the statistically significant results published in the literature and the outcome suggested a lack of evidential value. Fourth, a suite of sensitivity analyses were conducted to evaluate the robustness of these results, all of which converged on trivially small effect estimates. Overall, our results suggest the benefit of self-controlled practice on motor learning is small and not currently distinguishable from zero.

Keywords: Motor learning, retention, choice, “OPTIMAL” theory, meta-analysis, p-curve, publication bias

Meta-analytic findings of the self-controlled motor learning literature: Underpowered, biased, and lacking evidential value

Asking learners to control any aspect of their practice environment has come to be known as self-controlled practice in the motor learning literature (Sanli, Patterson, Bray, & Lee, 2013; Gabriele Wulf & Lewthwaite, 2016). The first published experiments to test self-controlled learning asked learners to control their augmented feedback schedule (Janelle, Barba, Frehlich, Tennant, & Cauraugh, 1997; Janelle, Kim, & Singer, 1995). For example, in an experiment by Janelle, Barba, Frehlich, Tennant, and Cauraugh (1997), participants practiced throwing tennis balls at a target with their non-dominant hand. The practice period occurred over two separate days. Participants were assigned to one of four experimental groups ($n = 12$): Self-controlled knowledge of performance, yoked-to-self-control, summary knowledge of performance after every five trials, and a knowledge of results only control group. The self-controlled group could request knowledge of performance whenever they wanted it, while each yoked group participant was matched with a self-control group counterpart and received knowledge of performance on the same schedule. The experimenter evaluated the participants' throws, identified the most critical error in their throwing form, and provided knowledge of performance via video feedback, along with directing attention to the error and giving prescriptive feedback. During a delayed-retention test, the accuracy, form, and speed of the throw were assessed. The results indicated that the self-control group threw more accurately and with better form than all other groups on the retention test. The self-control and yoked groups did not significantly differ in throwing speed, but the control group threw faster than the self-control group on the second retention block. The results were interpreted as evidence that the participants provided with choice were able to process information more efficiently than their counterparts who received a fixed schedule of feedback.

Figure 1 shows that the number of experiments comparing self-controlled groups to

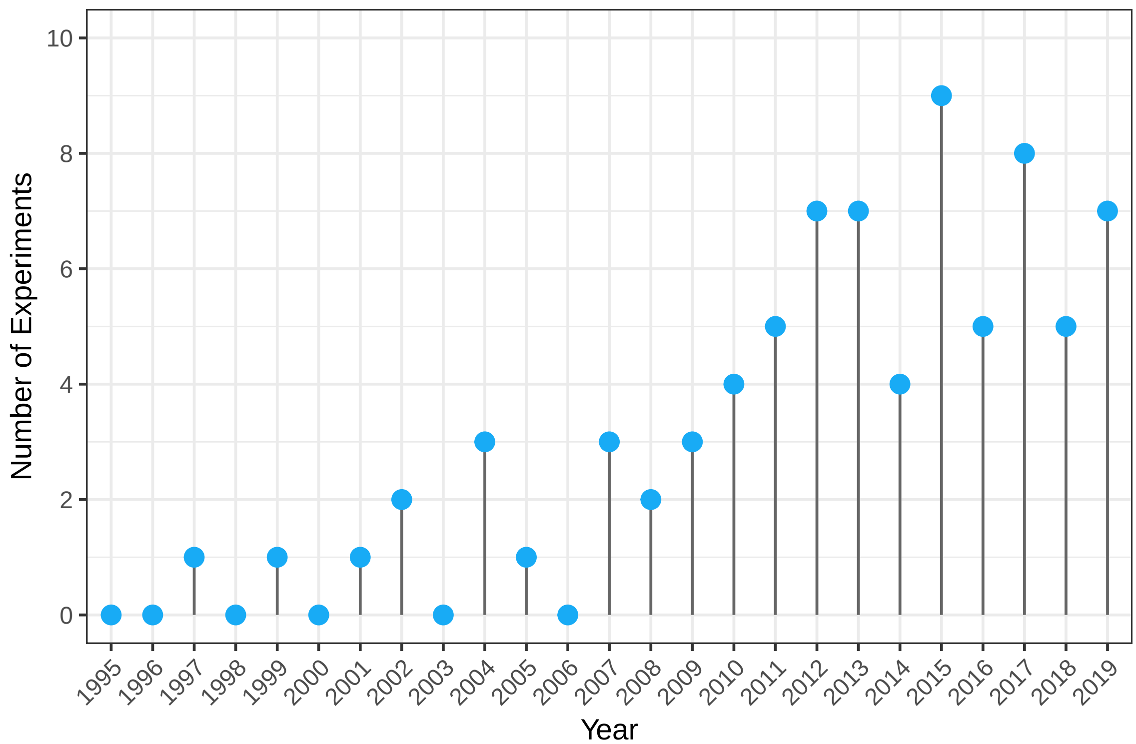

Figure 1

Number of self-controlled learning experiments meeting the inclusion criteria by year.

yoked groups has been increasing since the original experiments by Janelle and his colleagues (1997; 1995). Researchers have experimented with giving learners control over a variety of variables in the practice environment. A qualitative assessment of the literature suggests that self-control is generally beneficial regardless of choice-type (Gabriele Wulf & Lewthwaite, 2016). For example, self-control has been effective when participants have been provided choice over what can be considered instructionally-relevant variables, such as: knowledge of results (Patterson & Carter, 2010), knowledge of performance (Lim et al., 2015), concurrent feedback (Huet, Camachon, Fernandez, Jacobs, & Montagne, 2009), use of an assistive device (G. Wulf, Clauss, Shea, & Whitacre, 2001), observation of a skilled model (Lemos, Wulf, Lewthwaite, & Chiviacowsky, 2017), practice schedule (Will F. W. Wu & Magill, 2011), practice volume (Lessa & Chiviacowsky, 2015), and task difficulty (Leiker et al., 2016). Additionally, self-controlled benefits have also been found for instructionally-irrelevant variables, such as: the colour of various objects in the practice environment (Gabriele Wulf et al., 2018), other decorative choices (Iwatsuki, Navalta, & Wulf, 2019), and the choice of what to do after the retention test is complete (Lewthwaite, Chiviacowsky, Drews, & Wulf, 2015).

Despite the widespread optimism that self-controlled practice is useful for enhancing motor learning, researchers continue to debate the underlying mechanisms responsible for the effect (M. J. Carter & Ste-Marie, 2017b; Gabriele Wulf et al., 2018). Beginning with Janelle, Kim, and Singer (1995), both motivational and information processing mechanisms were proposed as possible explanations for self-control benefits. Researchers have since supported these two mechanisms and, from a motivational perspective, have posited that self-control enhances confidence (Chiviacowsky, Wulf, & Lewthwaite, 2012; Janelle, Kim, & Singer, 1995; Gabriele Wulf & Lewthwaite, 2016) and satisfies the basic psychological need for autonomy (Sanli, Patterson, Bray, & Lee, 2013; Gabriele Wulf & Lewthwaite, 2016), motivating motor performance and learning enhancement. Most self-controlled learning experiments, however, have involved participants making choices over potentially informative variables, which could act as a confounding variable. Citing this potential motivational/informational confound, Lewthwaite, Chiviacowsky, Drews, and Wulf (2015) experimented with providing instructionally-irrelevant choices, such as the colour of the golf balls to putt, the painting to hang on the wall, and what to do following the retention test. Lewthwaite and her colleagues reasoned that information processing explanations could not account for benefits due to these incidental choices, and instead motivational factors would be more likely. Consistent with the motivational hypothesis, participants exhibited significantly greater motor learning on a golf putting task (Experiment 1) and on a balance task (Experiment 2). Subsequently, several experiments have reported benefits with instructionally-irrelevant choices (Abdollahipour, Palomo Nieto, Psotta, & Wulf, 2017; Chua, Wulf, & Lewthwaite, 2018; Halperin, Chapman, Martin, Lewthwaite, & Wulf, 2017; Iwatsuki, Navalta, & Wulf, 2019; Gabriele Wulf, Chiviacowsky, & Cardozo, 2014; Gabriele Wulf et al., 2018), further reinforcing this motivational perspective.

A contrasting line of research has been reported by Carter and his colleagues (2014; 2017a, 2017b) in which informational factors, the second dominant perspective, are given more weight as an explanatory variable. In one experiment M. J. Carter, Carlsen, and Ste-Marie (2014), self-control participants were provided with choice over receiving knowledge of results, but divided into three experimental groups; those who could make their knowledge of results decision: before the trial, after the trial, or both (they would decide before, but could change their mind following the trial). Timing of the choice significantly attenuated the self-control benefit. While the self-after and self-both groups exhibited learning advantages relative to their yoked counterparts, the self-before group displayed no such advantage. The argument proffered by the researchers was that there was more informational value to be gained from knowledge of results requested after a trial than when it had to be requested before the outcome of the trial occurred (also see Chiviacowsky & Wulf, 2005).

In another experiment (M. J. Carter & Ste-Marie, 2017a), asking learners to complete an interpolated activity in the interval preceding their choice of whether to receive knowledge of results significantly attenuated the self-control benefit (also see Couvillion, Bass, & Fairbrother, 2020; Woodard & Fairbrother, 2020). As a final example, M. J. Carter and Ste-Marie (2017b) compared an instructionally-relevant choice group (i.e., when to receive knowledge of results) to an instructionally-irrelevant choice group (i.e., which video game to play after retention and which colour arm wrap to wear while practicing). Unlike the experiment by Wulf and colleagues (2018), Carter and Ste-Marie found that instructionally-relevant choices were more effective than task-irrelevant choices. Overall, they have used these different findings to tie self-controlled learning benefits to information-processing activities of the learner and, in particular, those related to the processing of intrinsic feedback (e.g., M. J. Carter & Ste-Marie, 2017a; Chiviacowsky & Wulf, 2005) and the provided knowledge of results (e.g., Grand et al., 2015).

In the present research, these different viewpoints concerning the mechanisms of self-controlled learning advantages were examined via meta-analysis with choice-type included as a moderator. The logic was that the motivational and informational perspectives would have different predictions. More specifically, from a motivation hypothesis, no moderating effect of choice-type on motor learning would be expected. In contrast, smaller effects for irrelevant-choice type, as compared to relevant-choice types would be expected from the information-processing perspective.

Beyond this interest in the possible theoretical mechanisms, a more important question addressed was whether there is in fact evidential value for the self-controlled learning benefit. This is of relevance because the current consensus in the field is that self-controlled practice is generally more effective than yoked practice (for reviews see Sanli, Patterson, Bray, & Lee, 2013; Ste-Marie, Carter, & Yantha, 2019; Gabriele Wulf & Lewthwaite, 2016). Reflecting this confidence in its benefits for motor learning, researchers have recommended adoption of self-control protocols in varied settings, such as medical training (Brydges, Carnahan, Safir, & Dubrowski, 2009; Jowett, LeBlanc, Xeroulis, MacRae, & Dubrowski, 2007; Gabriele Wulf, Shea, & Lewthwaite, 2010), physiotherapy (Hemayattalab, Arabameri, Pourazar, Ardakani, & Kashefi, 2013; Gabriele Wulf, 2007), music pedagogy (Gabriele Wulf & Mornell, 2008), strength and conditioning (Halperin, Wulf, Vigotsky, Schoenfeld, & Behm, 2018), and sports training (Janelle, Kim, & Singer, 1995; Sigrist, Rauter, Riener, & Wolf, 2013).

Problematic though is that recent, high-powered experiments with pre-registered analysis plans have failed to observe motor learning or performance benefits with self-control protocols (Grand, Daou, Lohse, & Miller, 2017; McKay & Ste-Marie, 2020; St. Germain et al., 2021; Yantha, 2019). Against the backdrop of the so-called replication crisis in psychology (Open Science Collaboration, 2015), there is reason for pause when evaluating the ostensible benefits of self-controlled learning. Further, Lohse, Buchanan, and Miller

(2016) have raised concerns about publication bias, uncorrected multiple comparisons, *p*-hacking, and other selection effects in the motor learning literature. Therefore, to address the impact of selection effects on estimates of the self-controlled learning effect, a weight function model (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019; Hedges & Vevea, 1996; McShane, Böckenholt, & Hansen, 2016; Vevea & Hedges, 1995; Vevea & Woods, 2005) with a one-tailed *p*-value cutpoint of .025 was fit to the dataset of effects to provide a pre-registered adjusted estimate of the overall self-controlled learning effect. Even the adjusted estimate is biased if the data generating processes are biased in ways not captured by the assumptions of the model, so further sensitivity analyses were conducted to estimate the average effect of self-control after correcting for selection effects (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019; Vevea & Woods, 2005). In parallel, in an effort to investigate the presence of evidential value in the literature, significant results were subjected to a *p*-curve analysis (Simonsohn, Nelson, & Simmons, 2014a; Simonsohn, Simmons, & Nelson, 2015). The *p*-curve analysis focuses exclusively on significant results and therefore is not affected by publication bias.

In sum, the objective of this meta-analysis was to estimate the true average effect of self-controlled learning and evaluate the evidential value of the self-controlled learning literature. Bias resulting from selective publication was addressed with weight function and *p*-curve models and effect size estimates were adjusted accordingly. A key theoretical question was also addressed through moderator analyses, but, to anticipate, inferences will depend on the reliability of the evidence overall. Finally, sensitivity analyses were conducted in addition to pre-registered analyses in an effort to understand the extent that our conclusions depended on the modelling techniques and assumptions adopted.

Methods

Pre-registration

The procedures followed to conduct this meta-analysis were pre-registered and can be viewed at https://osf.io/qbg69 (see Data, materials, and code availability section). This meta-analysis was retrospective and earlier samples of the literature had been meta-analyzed prior to this pre-registration, albeit with different data collection procedures, scope, and excluding recent experiments.

Literature search

The literature search and data extraction were conducted by three authors (BM, ZY, JH) and one research assistant (HS) independently. The goal of the search was to identify all articles that met the inclusion criteria for the meta-analysis. Specifically, randomized experiments were subject to five criteria for inclusion: 1) A self-control group in which participants were asked to make at least one choice during practice, 2) a yoked group that experienced the same practice conditions as the self-controlled group, 3) a delayed ~24-hour retention test or test with longer delay interval, 4) an objective measurement of motor performance, and 5) publication in a peer-reviewed journal or acceptance as part of a Master's or PhD thesis.

The search commenced at PubMed and Google Scholar with the following query: *self-control** OR *self-regulat** OR *self-direct** OR *learner-control** OR *learner-regulat** OR *learner-direct** OR *subject-control** OR *subject-regulat** OR *subject-direct** OR *performer-control** OR *performer-regulat** OR *performer-direct** AND *motor learning**. The query retrieved 9014 hits on PubMed and 98,600 hits on Google Scholar. Each researcher excluded hits based on title alone or title and abstract when necessary, and quit searching the databases at self-selected intervals following extended periods of excluding 100% of search results. Following an initial run of searching databases, each researcher employed their own search strategies, including reviewing the reference sections of reviews and included articles, consulting the "OPTIMAL' ' theory website[1], and searching the ProQuest Thesis database.

This literature search process resulted in 160 articles that could not be excluded without consulting the full-text of the article. All 160 articles were coded for inclusion or exclusion by two researchers independently. All instances of disagreement between coders were reviewed by three authors (BM, ZY, and JH), and consensus was reached in each case. Disagreements were infrequent and were often caused by a lack of clarity in the articles (e.g., 100% knowledge of results groups labelled as yoked groups). None of the coding disagreements evolved into conceptual disagreements. Rather, in each case, it was identified that one coder had missed a detail in the full text that changed its inclusion eligibility. Subsequent to this process, a total of 73 articles, which included 78 experiments, met the inclusion criteria (see Table 1).

Dependent variable selection

The focus of this meta-analysis was on performance outcomes associated with the goal of the skill. The primary theoretical perspectives offered as an account for self-controlled learning are likewise focused on performance outcomes. For example, the "OPTIMAL' ' theory proposes that a learner's movements become coupled with the goal they are trying to achieve when they experience autonomy-support during practice (Gabriele Wulf & Lewthwaite, 2016). To reflect this focus, a dependent measure priority list was developed that gave higher priority to absolute error measures and less priority to consistency measures, time/work measures, and form scores. Dependent measure priority was ordered as follows: 1) absolute error (and analogous measures: radial error, points in an accuracy measure), 2) root-mean-square-error (RMSE), 3) absolute constant error, 4) variable error, 5), movement

[1] The webpage link that was consulted (https://optimalmotorlearning.com/index.php/did-you-know-that/) is no longer available. A new webpage devoted to "OPTIMAL" theory can be accessed using the following link: https://gwulf.faculty.unlv.edu/optimal-motor-learning/.

time (and distance travelled), 6) movement form – expert raters, 7) otherwise unspecified objective performance measure reported first in research report.[2] In the event that multiple measures of motor performance were reported for an experiment, effect sizes were calculated for the highest priority measure reported in the study. In experiments with multiple self-control groups and one yoked group, the self-control groups were combined (Higgins & Green, 2011). If multiple choice-types or sub-populations were included in an experiment, combined and individual effects were calculated for inclusion in moderator analyses.

Many of the self-controlled learning experiments analyzed in this study included multiple dependent measures. However, including multiple measures from the same experiment introduces bias and inflates Type 1 error (Scammacca, Roberts, & Stuebing, 2014). Although there are a variety of methods for dealing with multiple measures from the same studies in meta-analysis, we chose to create a priority list and always selected the highest priority dependent measure that was reported. If the highest priority measure was not described in adequate detail to calculate the effect size, the authors were contacted and the data were requested. If the authors could not provide the data for the highest priority dependent measure reported in their study, the experiment was left out of our analysis.

The rationale for selecting the approach we did was based on five considerations. First, our interest was in motor learning as reflected by an enhanced capability to perform a skill. Motor learning studies often report multiple error measures, but they are not equally coupled with performance outcome. Constant error, for example, was not included on the priority list because it is possible to have zero constant error while performing terribly overall. Therefore, we chose to prioritize measures that could be considered to be tightly coupled with performance, like absolute error, RMSE, and absolute constant error. If these

[2] Radial error, accuracy points, and distance travelled were added to the pre-registered dependent measures as they arose during data-extraction. Decisions were made blind to the data by an author not involved in said extraction (BM or DSM).

measures were not used, measures that are only correlated with performance, such as variable error, movement time, and movement form, were selected. We reasoned this selection strategy would focus the analysis on measures related to improved skill while de-emphasizing other effects. Second, we reasoned that averaging across dependent measures could introduce additional heterogeneity to the analysis by including potentially disparate dependent measures. The third, fourth, and fifth considerations all relate to avoiding bias but differ with regard to the source of the bias and the alternate method that would include such bias. Thus, the third consideration was that imposing a priority list was thought to better avoid biases that could emerge from selecting the most focal measure in a given study, because an unknowable percentage of studies may have defined the focal measure based on the strength of the findings. Fourth, we reasoned that some measures may only get reported if they support the predicted benefit of self-control. Scammacca, Roberts, and Stuebing (2014) reported that effect size estimates were inflated when random dependent measures were selected in a meta-analysis case study, perhaps reflecting a selective reporting bias. Averaging across all reported measures–a fair alternative to our approach–could conceivably pick up some of this reporting bias. Fifth, we ignored lower priority measures with data when higher priority measures lacked data because we reasoned there could be a systematic reason for this pattern: preference for reporting data associated with positive effects. Indeed, there were articles where the only measure reported with sufficient data to calculate an effect size was also the only measure with a significant result (e.g., Gabriele Wulf, Raupach, & Pfeiffer, 2005).

Table 1

Experiment characteristics and moderator coding.

Authors	Year	Setting	Compensation	Choice-type	Population	Retention	N	Published
Aiken et al.	2012	Applied	Not stated	Observation	Adult	24-hr	28	Yes
Alami	2013	Lab	Yes	Feedback (KR)	Adult	24-hr	22	No
Ali et al.	2012	Lab	Not stated	Feedback (KR)	Adult	24-hr	48	Yes
Andrieux et al.	2016	Lab	Not stated	Task difficulty	Adult	24-hr	48	Yes
Andrieux et al.	2012	Lab	Not stated	Task difficulty	Adult	24-hr	38	Yes
Arsal	2004, Exp 1	Lab	Not stated	Feedback (KR)	Adult	48-hr	28	No
Arsal	2004, Exp 2	Lab	Not stated	Feedback (KR)	Adult	48-hr	28	No
Barros	2010, Blocked	Lab	Not stated	Feedback (KR)	Adult	24-hr	48	No
Barros	2010, Random	Lab	Not stated	Feedback (KR)	Adult	24-hr	48	No
Barros et al.	2019, Exp 1	Lab-Applied	No	Feedback (KR)	Adult	24-hr	60	Yes
Barros et al.	2019, Exp 2	Lab	No	Feedback (KR)	Adult	24-hr	60	Yes
Bass	2015	Lab	No	Feedback (KR)	Adult	24-hr	20	No
Bass	2018	Applied	No	Feedback (KR)	Adult	24-hr	60	No
Brydges et al.	2009	Applied	Not stated	Observation	Adult	>48-hr	48	Yes
Bund & Weimeyer	2004	Lab-Applied	No	Observation	Adult	24-hr	52	Yes

Carter & Patterson	2012	Lab	Not stated	Feedback (KR)	Adult	24-hr	20	Yes
Carter & Patterson	2012	Lab	Not stated	Feedback (KR)	Older	24-hr	20	Yes
Carter & Patterson	2012	Lab	Not stated	Feedback (KR)	Two	24-hr	40	Yes
Chen et al.	2002	Lab	Yes	Feedback (KR)	Adult	48-hr	48	Yes
Chiviacowsky	2014	Lab	Not stated	Feedback (KR)	Adult	24-hr	28	Yes
Chiviacowsky & Lessa	2017	Lab	Not stated	Feedback (KR)	Oider	48-hr	22	Yes
Chiviacowsky & Wulf	2002	Lab	Not stated	Feedback (KR)	Adult	24-hr	30	Yes
Chiviacowsky et al.	2012	Lab	Not stated	Feedback (KR)	Clinical	24-hr	30	Yes
Chiviacowsky et al.	2008	Lab	Not stated	Feedback (KR)	Children	24-hr	26	Yes
Chiviacowsky et al.	2012	Lab	Not stated	Assistive device	Clinical	24-hr	28	Yes
Davis	2009	Applied	Not stated	Model	Adult	24-hr	24	No
Fagundes et al.	2013	Lab-Applied	Not stated	Feedback (KR)	Adult	48-hr	52	Yes
Fairbrother et al.	2012	Lab	Not stated	Feedback (KR)	Adult	24-hr	48	Yes
Ferreira et al.	2019	Lab	Not stated	Feedback (KR)	Adult	24-hr	60	Yes
Figueiredo et al.	2018	Lab	No	Feedback (KR)	Adult	24-hr	30	Yes
Ghorbani	2019, Exp 2	Lab-Applied	Not stated	Feedback (KR)	Adult	24-hr	36	Yes
Grand et al.	2015	Lab	No	Feedback (KR)	Adult	24-hr	36	Yes
Grand et al.	2017	Lab	Yes	Incidental	Adult	>48-hr	68	Yes

Hansen et al.	2011	Lab	No	Feedback (KR)	Adult	24-hr	24	Yes
Hartman	2007	Lab	Not stated	Assistive device	Adult	24-hr	18	Yes
Hemayettalab et al.	2013	Lab	Not stated	Feedback (KR)	Clinical	24-hr	20	Yes
Ho	2016	Lab	Not stated	Amount of practice	Adult	24-hr	120	No
Holmberg	2013	Lab-Applied	No	Feedback (KP)	Adult	24-hr	24	No
Huet et al.	2009	Lab-Applied	Not stated	Feedback (Concurrent)	Adult	24-hr	20	Yes
Ikudome et al.	2019, Exp 1	Lab-Applied	No	Incidental	Adult	24-hr	40	Yes
Ikudome et al.	2019, Exp 2	Lab-Applied	No	Observation	Adult	24-hr	40	Yes
Jalalvan et al.	2019	Lab-Applied	Not stated	Task difficulty	Adult	24-hr	60	Yes
Janelle et al.	1997	Lab-Applied	Yes	Feedback (KP)	Adult	>48-hr	48	Yes
Jones	2010	Lab	Yes	Repetition schedule	Adult	24-hr	40	No
Kaefer et al.	2014	Lab	No	Feedback (KR)	Adult	24-hr	56	Yes
Keetch & Lee	2007	Lab	Yes	Repetition schedule	Adult	24-hr	96	Yes
Kim et al.	2019	Lab	Yes	Feedback (KR)	Adult	24-hr	42	Yes
Leiker et al.	2016	Lab-Applied	Not stated	Task difficulty	Adult	>48-hr	60	Yes
Leiker et al.	2019	Lab	Not stated	Task difficulty	Adult	>48-hr	60	Yes
Lemos et al.	2017	Applied	No	Observation	Children	24-hr	24	Yes
Lessa & Chiviacowsky	2015	Applied	Not stated	Amount of practice	Older	48-hr	36	Yes

Lewthwaite et al.	2015, Exp 1	Lab-Applied	Not stated	Incidental	Adult	24-hr	24	Yes
Lewthwaite et al.	2015, Exp 2	Lab	Not stated	Incidental	Adult	24-hr	30	Yes
Lim et al.	2015	Applied	Not stated	Feedback (KP)	Adult	24-hr	24	Yes
Marques & Correa	2016	Applied	Not stated	Feedback (KP)	Adult	48-hr	70	Yes
Marques et al.	2017	Applied	Not stated	Feedback (KP)	Adult	24-hr	30	Yes
Norouzi et al.	2016	Lab	Not stated	Feedback (KR)	Adult	24-hr	45	Yes
Nunes et al.	2019	Lab-Applied	No	Feedback (KP)	Older	24-hr	40	Yes
Ostrowski	2015	Lab	Not stated	Feedback (KR)	Adult	24-hr	80	No
Patterson & Carter	2010	Lab	Yes	Feedback (KR)	Adult	24-hr	24	Yes
Patterson & Lee	2010	Lab-Applied	Yes	Task difficulty	Adult	48-hr	48	Yes
Patterson et al.	2013	Lab	Yes	Feedback (KR)	Adult	24-hr	48	Yes
Patterson et al.	2011	Lab	Yes	Feedback (KR)	Adult	24-hr	60	Yes
Post et al.	2016	Lab-Applied	No	Feedback (KP)	Adult	24-hr	44	Yes
Post et al.	2011	Applied	No	Amount of practice	Adult	24-hr	24	Yes
Post et al.	2014	Applied	Not stated	Amount of practice	Adult	24-hr	30	Yes
Rydberg	2011	Applied	Not stated	Repetition schedule	Adult	24-hr	16	No
Sanli & Patterson	2013	Lab	No	Repetition schedule	Adult	24-hr	24	Yes
Sanli & Patterson	2013	Lab	No	Repetition schedule	Children	24-hr	24	Yes

Ste-Marie et al.	2013	Applied	No	Feedback (KP)	Children	24-hr	60	Yes
Tsai & Jwo	2015	Lab	Yes	Feedback (KR)	Adult	24-hr	36	Yes
von Lindern	2017	Lab	Not stated	Feedback (KR)	Adult	24-hr	48	No
Williams et al.	2017	Lab	Yes	Feedback (Concurrent)	Adult	24-hr	29	Yes
Wu & Magill	2011	Lab	No	Repetition schedule	Adult	24-hr	30	Yes
Wu	2007, Exp 1	Lab-Applied	Yes	Repetition schedule	Adult	24-hr	30	No
Wulf & Adams	2014	Lab	No	Repetition schedule	Adult	24-hr	20	Yes
Wulf &Toole	1999	Lab-Applied	Yes	Assistive device	Adult	24-hr	26	Yes
Wulf et al.	2015, Exp 1	Lab-Applied	No	Repetition schedule	Adult	24-hr	68	Yes
Wulf et al.	2001	Lab-Applied	Yes	Assistive device	Adult	24-hr	26	Yes
Wulf et al.	2018, Exp 1	Lab-Applied	No	Incidental	Adult	24-hr	32	Yes
Wulf et al.	2018, Exp 2	Lab-Applied	No	Incidental	Adult	48-hr	28	Yes
Wulf et al.	2018, Exp 2	Lab-Applied	No	Observation	Adult	48-hr	28	Yes
Wulf et al.	2018, Exp 2	Lab-Applied	No	Two	Adult	48-hr	42	Yes
Wulf et al.	2005	Applied	No	Observation	Adult	>48-hr	26	Yes

Note. KR = Knowledge of results; KP = Knowledge of performance.

Data extraction

The four researchers separated into pairs and half of the included experiments were coded independently by one pair. The other half were coded independently by the other pair. The coding included varied moderators, publication year, and sample size. Also Hedges' g was calculated from reported statistics and sample size using the `compute.es` package (Re, 2013) in `R` (R Core Team, 2020). Effect sizes were calculated from means and standard deviations, test statistics like t and F, or from precisely reported p-values. When covariates were included in the analysis, the correlation coefficient for the covariate - dependent measure relationship was required to calculate accurate effect sizes. Since this information is often not reported, authors were contacted and the information was requested. One effect size was calculated for each of three time points for each experiment: Acquisition, retention, and transfer.

The independent data extractions were compared and inconsistent results were highlighted. There was 89% absolute agreement between pairs of coders on 1344 data points. For those with disagreement, one of the researchers from the other coding pair reviewed the relevant experiment to confirm the value to be used in the analysis.[3]

Several articles failed to report the data necessary to calculate effect sizes at some or all time-points. A total of 39 authors were emailed with requests for missing data and 17 were able to provide data following a minimum one month period following the request. After requesting missing data, 25 experiments were excluded from primary analyses for missing retention data. A total of 52 effects from 51 experiments reported in 46 articles were

[3] On one occasion, the third researcher was unable to match either effect calculation, so the involved researchers discussed the issue, determined the source of the inconsistency, and asked a fourth researcher to recalculate the effect size with clear instructions for avoiding confusion. The source of inconsistency was simply a rounding error when combining multiple groups and the fourth researcher was able to corroborate the calculation.

included in the primary meta-analysis.

In addition to extracting effect sizes, inferential statistics were scraped from published experiments that reported a statistically significant effect at retention. Two authors (BM and JH) independently completed a *p*-curve disclosure form consisting of a direct quote of the stated hypotheses for each experiment, the experimental design, and a direct quote of the results indicating a significant result **(see Appendix A)**. There was 94% absolute agreement between the independent forms. Mismatches were resolved with consensus.

Outlier screening

The meta-analysis `R` package `metafor` (Viechtbauer, 2010) was used to screen the data for potentially influential outliers (see analysis script). In order to identify outlier values and exclude them from further analyses, the following nine influence statistics were calculated: a) externally standardized residuals, b) DFFITS values, c) Cook's distances, d) covariance ratios, e) DFBETAS values, f) the estimates of t^2 when each study is removed in turn, g) the test statistics for (residual) heterogeneity when each study is removed in turn, h) the diagonal elements of the hat matrix, and i) the weights (in %) given to the observed outcomes during the model fitting. Any experiment with effects identified as extremely influential by any three of the influence metrics were removed from subsequent analyses.

Pre-specified analyses

Random effects model

A naïve random effects model was fit to the retention effect sizes to estimate the average reported effect of self-controlled learning and to assess heterogeneity in effect sizes between experiments. Heterogeneity was evaluated with the Q statistic and described with I^2. A mixed-effects model was fit to evaluate whether differences in experimental design or sample characteristics moderated the effect of self-controlled learning.

Moderator analyses

Moderators were determined based on the authors' collective knowledge of the self-controlled learning literature. We coded for discrete differences in protocols between experiments to investigate whether differing methodologies resulted in different effect size estimates. Further, based on a meta-analysis reporting that the effect of choice on intrinsic motivation can be moderated by whether participants were compensated for completing the study (Patall, Cooper, & Robinson, 2008), we also coded for compensation type. Finally, we investigated whether publication status was a moderator of the effect of self-control as part of our overall approach to examining the impact of publication bias on the self-controlled learning literature. The following six moderators were analyzed separately in mixed-effects models: a) *Choice-type*: Choices were categorized as either instructionally-irrelevant, knowledge of results, knowledge of performance, concurrent feedback, amount of practice, use of assistive device, practice schedule, observational practice, or difficulty of practice; b) *Experimental setting*: Experiments were categorized as either laboratory, applied, or laboratory-applied. We defined a laboratory setting as one where learners are asked to acquire a skill not typically performed in everyday life. We defined an applied setting as one where learners are asked to acquire a skill often performed outside of a laboratory. Finally, we defined a laboratory-applied setting as one where learners are asked to acquire a skill resembling skills often performed outside the laboratory but with researcher-contrived differences; c) *Sub-population*: The following subgroups were analyzed: Adult (18-50 years of age), children/adolescents (under 18-years old), older adult (over 50-years-old), and clinical (clinical population defined by the research article); d) *Publication status*: Articles were classified as published or unpublished (e.g., theses); e) *Compensation*: Whether participants were compensated for participating in the experiment was categorized as compensated, not compensated, or not stated; f) *Retention delay-interval*: Coded as 24-hour, 48-hours, or >48-hours.

Adjusting for selection effects

Selection bias in the motor learning literature is likely caused by filtering based on the statistical significance of results (Lohse, Buchanan, & Miller, 2016). To assess and adjust for selection effects, the `R` package `weightr` (Coburn & Vevea, 2017) was used to fit a Vevea-Hedges weight function model to the retention data (Vevea & Hedges, 1995). The weight-function model estimates the true average effect, heterogeneity, and the probability that a non-significant result survives censorship and is available for analysis. Selection effects are modelled by a step function that divides the effects into two bins at one-tailed p =.025, coinciding with a two-tailed p-value of .05. The probability of a non-significant effect surviving censorship to appear in the model is estimated relative to the probability of observing a study with a significant effect. The selection-adjusted model was compared to the naïve random effects model with a likelihood ratio test. Better fit from an adjusted model suggests selection bias in the literature.

The adjusted estimate from the weight-function model was pre-registered as the primary estimate of the true average effect in this meta-analysis. Please note that while the weight-function model attempts to estimate the true effect of self-controlled learning after correcting for selection biases, the estimated effect cannot be considered definitive. Nevertheless, the adjusted estimate is likely less biased than the naïve random effects estimate (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019; Hong & Reed, 2021; Kvarven, Strømland, & Johannesson, 2020; Vevea & Hedges, 1995). The difference between the estimates can be informative about the potential impact of selection biases, with larger disparities between models suggesting greater selection effects.

P-curve analysis

To investigate the evidential value of the self-controlled learning literature, the significant positive results at retention reported in peer-reviewed journals were submitted to a p-curve analysis (Simonsohn, Simmons, & Nelson, 2015). To be included in the analysis, articles needed to meet the following criteria: a) be a published article; b) state explicitly that self-controlled learning was expected to be more effective than yoked practice; c) report inferential statistics comparing a self-control group and a yoked group directly on a retention test; d) conclude that the self-control group performed significantly better than the yoked group. If the article included multiple dependent measures showing a significant effect, the dependent measure priority list was used to select the highest priority measure. If only one measure was reported as significant, that effect was included even if the experiment included higher priority measures that were null. This resulted in a slightly different sample of effects from the random effects and weight-function models.

The distribution of significant *p*-values is a function of the power of the experiments included in the analysis. If a *p*-curve included only Type 1 errors, the expected distribution would be uniform. As the power of included experiments increases, so too does the amount of right skew in the *p*-curve, with smaller *p*-values appearing more frequently than large *p*-values. The *p*-curve analysis tests the null hypothesis that there is no evidentiary value by analyzing the amount of right skew in the distribution of *p*-values. Conversely, if researchers peek at their data and stop collecting when they reach statistical significance, a practice known as *p*-hacking, the distribution of significant *p*-values under the null would be left skewed, with *p*-values near .05 occurring more frequently. Varying mixtures of true effect sizes and intensities of *p*-hacking produce varying shapes of *p*-curve, therefore the observed *p*-curve was compared to the distribution of *p*-values expected if the studies were conducted with 33% power. It is unlikely that researchers would continuously conduct experiments that fail >66% of the time whilst studying the self-controlled learning phenomenon. Observing a *p*-curve significantly "flatter'' than what would be expected with 33% power would suggest a lack of evidential value among the significant results (Simonsohn, Nelson, & Simmons, 2014b, 2014a).

Sensitivity analyses

The primary analyses were followed up with several sensitivity analyses. Sensitivity analyses are used to evaluate the sensitivity of the results to the specific parameters chosen for the original analyses. The self-controlled learning literature, like many areas of behavioural research, was not produced exclusively by registered experiments with pre-specified analysis plans and 100% reporting frequency. The complexity of selection effects at various levels, including editorial decisions, author decisions, analysis decisions, and missing data, renders the accuracy of modeled effects impossible to estimate (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019). Producing a range of estimates based on varying assumptions is intended to provide the reader with a broader picture of the uncertainty of the point estimates in the primary analyses.

Bias correction methods vary in their performance depending on the total amount of heterogeneity, the true average effect size, the amount of publication bias, and the intensity of *p*-hacking in the data (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019). To determine which bias correction models perform well in the various plausible conditions for data in this meta-analysis, model performance checks were conducted using the Meta-Showdown Explorer shiny app developed by Carter and colleagues (2019). Simulated conditions were as follows: Medium publications bias (significant results published at 100% frequency, non-significant published at 20% frequency, wrong direction effects published at 5% frequency), medium questionable research practice environment (QRP; see Carter et al. (2019) for detailed explanation of QRP environment), $\tau = 0, .2$; $g = 0, .2, .5$; $k = 60$, good performance defined as a maximum of .1 upward or downward bias, and maximum mean absolute error of .1, also tested with maximum bias and error values of .15. With good performance defined by a maximum bias in either direction of .1 and maximum absolute error of .1, the weight function model and, to a lesser extent, *p*-curve models provided coverage across all plausible conditions except the highest heterogeneity condition ($\tau = .4$).

With good performance defined as a maximum bias and error of .15, the precision-effect with standard error (PEESE) method provided good performance in all conditions. Therefore, sensitivity analyses were conducted on effect size data via *p*-curve and PEESE methods. An additional sensitivity analysis of the estimated power among included studies was conducted with the *z*-curve (Bartoš & Schimmack, 2020). *Z*-curve, like *p*-curve, analyzes only statistically signficant results and estimates the power of the included studies (called expected replication rate, ERR). However, unlike *p*-curve, *z*-curve is robust to heterogeneity because it fits a finite mixture model of seven distributions, allowing the underlying true effects to vary. Further, *z*-curve also estimates the power of all studies that have been conducted (called expected discovery rate, EDR) which can be compared to the observed discovery rate in order to test for the presence of publication bias.

Primary *p*-curve

A leave-one-out analysis of *p*-curve results was conducted to assess the extent to which the primary results depended on the inclusion of one or two extreme results. Results that depend on the inclusion of one or two extreme results should not be considered robust.

Results

Outlier removal

Two studies were flagged as significantly influential outliers by all nine influence metrics calculated during data screening: Lemos, Wulf, Lewthwaite, and Chiviacowsky (2017), $g = 3.7$, and Marques, Thon, Espanhol, Tani, and Corrêa (2017), $g = 3.95$. No other effect sizes were identified as outliers by any metric. Both outliers were removed from all subsequent analyses.

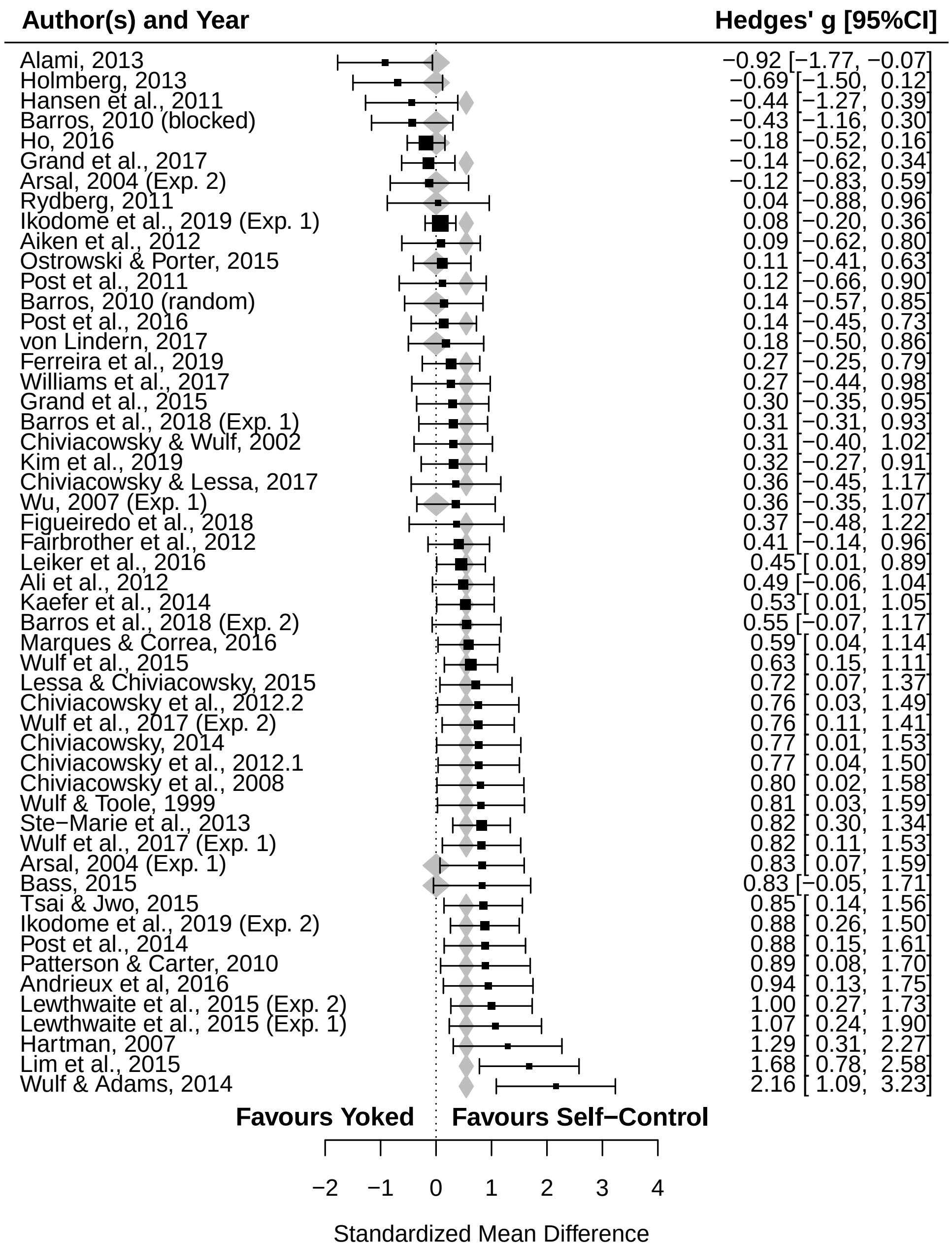

Figure 2

Forest plot of Hedges' g (95% CI) for self-controlled versus yoked groups on retention tests. Size of squares is proportional to $1/\sigma^2$ *(precision). Light grey polygons represent 95% CI estimates from publication-status moderator analysis. Estimates from unpublished studies center on g = .003 and published studies on g = .54*

Naïve random effects model

The naïve random effects model estimated the average treatment effect of self-controlled practice, $g = .44$ ($k = 52$, $N = 3134$, 95% CI [.31, .56]). However, there was significant variability in the average effect estimated across experiments, $Q(df = 51) = 103.45$, $p < .0001$, $\tau = .31$. It was estimated that 47.9% (I^2) of the total variability in effect sizes across experiments was due to true heterogeneity in the underlying effects measured (see Figure 2).

Moderator analyses

Six moderators selected for theoretical and/or methodological reasons were tested separately. Five moderators failed to account for a significant amount of heterogeneity: experimental setting ($p = .46$, $R^2 = 1\%$), compensation ($p = .99$, $R^2 = 0\%$), choice-type ($p = .71$, $R^2 = 0\%$), sub-population ($p = .74$, $R^2 = 0\%$), and retention interval ($p = .54$, $R^2 = 0\%$). One moderator, publication status, accounted for a statistically significant amount of heterogeneity, $p < .0001$, $R^2 = 48\%$. Among published experiments self-controlled practice had a strong benefit, $g = .54$, 95% CI [.28, .81]. However, among unpublished experiments self-controlled practice had essentially no effect, $g = .003$, 95% CI [-.23, 24].

Selection model

The weight-function model combines an effect size model and a selection model (Hedges & Vevea, 1996). The effect size model is equivalent to the naïve random effects model, specifying what the distribution of effect sizes would be in the absence of publication bias or other selection effects. The selection model accounts for the probability a given study survives selection based on its p-value and specifies how the effect size distribution is modified by selection. A weight-function model with a p-value cutpoint of (one-tailed) .025 was fit to the retention effect size estimates. The results of a likelihood ratio test suggest the adjusted model was a significantly better fit to the data than the unadjusted model, $\chi^2(df =$

1) = 21.18, $p < .0001$.[4] The adjusted effect size estimate was significantly different from zero, $g = .107$, $p < .001$, 95% CI [.05, .17]. According to the adjusted model, non-significant results were 6% as likely to survive selection as significant results (see Figure 3). Note that the `weightr` function failed to estimate the random effects model and the results reported here are based on a fixed-effect estimate.

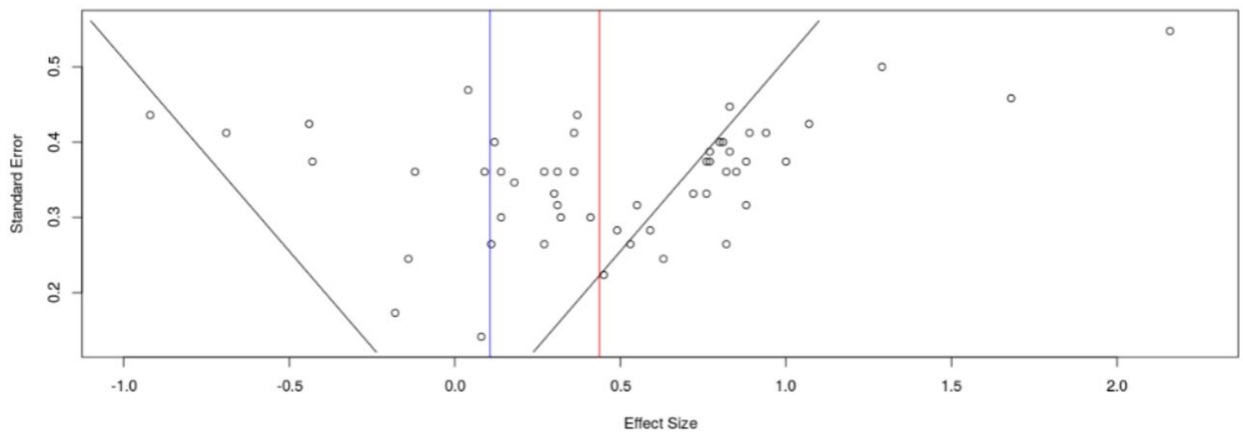

Figure 3

Funnel plot of self-controlled learning studies at retention. Standard error is plotted on the y-axis and Hedges' g is plotted on the x-axis. Red line represents naïve random effects model estimate (g = .44); blue line represents weight-function model adjusted estimate (g = .11); black contour lines represent two-tailed p-values of .05. In the absence of bias (and other forms of heterogeneity), the most precise experiments would centre on the naïve random effects estimate near the bottom of the plot and as experiments get progressively less precise they would move up the plot and spread out symmetrically. In the presence of bias, one would expect experiments to track to the right side of the positive contour line. The clustering of experiments just to the right of the positive contour line in the above plot suggests substantial bias.

[4] Be aware that the likelihood ratio test is not robust to misspecification of the random effects model (Hedges & Vevea, 1996).

P-curve

The purpose of the *p*-curve analysis was to investigate the evidential value in the published reports ($N = 26$) of statistically significant self-controlled learning benefits. Visual inspection of Figure 4 reveals a v-shaped distribution with the greatest frequency of *p*-values in the <.05 bin. The observed *p*-curve was significantly flatter than would be expected if the experiments had 33% power, $p = .0035$, indicating an absence of evidential value. Conversely, the half *p*-curve (Simonsohn, Simmons, & Nelson, 2015) was significantly right skewed, suggesting the presence of evidential value. Sensitivity analysis, however, revealed that the half curve does not remain significantly right skewed following removal of the most extreme *p*-value from the sample. The estimated power of the included studies was 5%, 95% CI [5%, 17%].

Interim discussion

The primary results described above suggest that selection effects have caused a seriously distorted record of self-controlled learning. Estimated benefits are less than one third of the naïve estimate, $g = .107$, 95% CI [.05, .17]. The *p*-curve analysis failed to detect robust evidence of a self-controlled learning effect. The performance of the weight-function model depends on the specific conditions present in the meta-analysis, although these conditions are unknowable (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019). It was necessary to conduct sensitivity analyses with additional bias correction methods to assess the reliability of the selection-adjusted weight-function model estimate. Based on performance checks conducted under a range of plausible conditions, it was determined that sensitivity analyses conducted with a PEESE meta-regression and *p*-curve effect size estimation would provide good performance coverage across most plausible conditions.

Sensitivity analyses

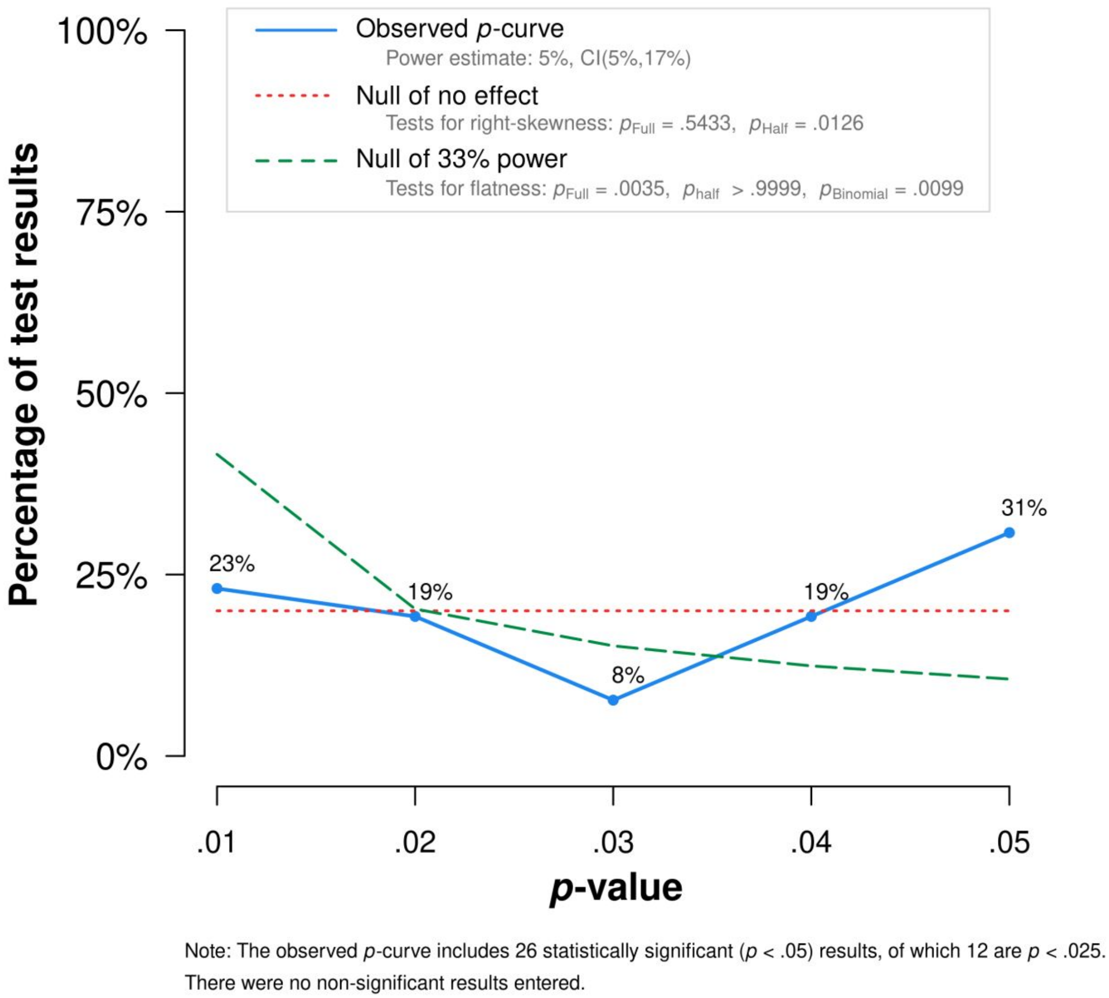

Figure 4

P-*curve analysis of published experiments that were statistically significant at retention. If the included experiments are studying a true null hypothesis the expected distribution of p-values is uniform, represented by the dotted line. If the experiments are studying a true effect, the expected distribution becomes increasingly right skewed as a function of statistical power. The expected right skewed distribution associated with 33% power is plotted by the dashed line. The observed p-curve is plotted by the solid line and was substantially flatter than the 33% power distribution. The half p-curve analysis included p-values below $p = .025$ and was significantly right skewed. The right skew did not survive deletion of the most extreme value.*

Precision-effect with standard error (PEESE) model

When publication bias is present in a body of evidence, sample size and effect size can be negatively correlated (Stanley & Doucouliagos, 2014). The PEESE model fits a quadratic relationship between effect size and standard error to reflect the intuition that publication bias is stronger for low precision studies than high precision studies. The rationale is that low precision studies need to overestimate effects to achieve significance and get published, while high precision studies can publish without exaggerated effects; thus, creating greater publication bias among lower precision studies (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019; Stanley & Doucouliagos, 2014). A weighted-least-squares regression model was fit with effect size regressed on the square of the standard error, weighted by the inverse of the variance:

$$g_i = b_0 + b_1 se_i^2 + e_i$$

The PEESE method estimated a non-significant benefit of self-controlled learning after controlling for publication bias, $g = .054$, $p = .659$.

P-curve effect estimation

A *p*-curve model was fit to the overall retention effect size data, unlike the first primary *p*-curve which was fit to the reported significant results. The *p*-curve is a function of sample size and effect size, and because sample size is known, the effect size that provides the best fit to the observed *p*-curve can be estimated (Simonsohn, Nelson, & Simmons, 2014b). A *p*-curve analysis conducted with thR package `dmetar` (Harrer, Cuijpers, Furukawa, & Ebert, 2019) was used to estimate the average effect size among the statistically significant effects in the meta-analysis. The model estimated an average effect of $g = .035$.[5] The estimated power of included studies was 7%, 95% *CI* [5%, 22%]. Unfortunately, *p*-curve does not perform well in the presence of heterogeneity and these results should be interpreted cautiously.

[5] The *p*-curve of effect sizes was significantly flatter than the expected 33% power curve as well, $p = .009$.

Z-curve

A z-curve was fit to the overall retention data and estimated the power of statistically significant studies (ERR) as 14%, 95% CI [5%, 35%]. The power of all studies conducted (EDR) was estimated as 6%, 95% CI [5%, 14%]. The 95% confidence intervals for both the ERR and EDR failed to include the observed discovery rate of 48%, suggesting significant publication bias in the data.

Acquisition and transfer

In light of the evidence that experiments are apparently selected for positive self-controlled learning effects at retention, pre-planned exploratory estimates of the effect of self-controlled practice on acquisition and transfer performance can no longer be considered reliable. However, given that some have argued that transfer tests are more senstive measures of motor learning than delayed retention tests (Chiviacowsky & Wulf, 2002; Fairbrother, Laughlin, & Nguyen, 2012), the transfer test data were analzed via both naïve random effects and weight function models. The naïve estimate at transfer was $g = .52$, while the bias corrected estimate was $g = .17$, $p = .24$. As with delayed retention, the selection model provided a better fit to the transfer data than the naïve model, $p = .008$. The primary take away from these analyses is that the reported self-controlled learning effects to date are unreliable.

Discussion

The primary objective of this meta-analysis was to assess the effect of providing choices during the acquisition of a motor skill on delayed retention performance in the general population. A secondary objective was to test between motivation and informational explanations for self-controlled learning benefits by investigating whether choice-type moderates the effect of choice. To this aim, an extensive search for experiments that compared self-controlled practice to a yoked comparison group was conducted. Effect size and moderator data were ascertained from data reported in the research articles or, in some cases, received directly from the authors of the studies. Efforts were taken to ensure that each effect size calculation and moderator code could be reproduced by an independent party. In parallel, the results of published experiments that achieved a hypothesized statistically significant result in favour of self-control were extracted directly from the articles and outlined in a *p*-curve disclosure form **(see Appendix A)**. Pre-registered primary analyses were applied to the data and results were followed up with a suite of sensitivity analyses.

The naïve random effects model estimated a benefit from self-controlled practice of $g = .44$. However, the naïve model fails to account for selection effects, such as publication bias and *p*-hacking, and as such overestimates the true average effect when these selection effects are present (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019; Hedges & Vevea, 1996; Stanley & Doucouliagos, 2014). Publication status was a significant moderator of the self-controlled practice effect, accounting for 48% of the total heterogeneity in the model. Published experiments reported an average benefit of $g = .54$ while unpublished experiments reported no benefit at all on average. It is possible that researchers use statistical significance, typically defined as $p < .05$ on a two-tailed test, to filter their results for publication. To account for potential selection effects driven by statistical significance, a weight-function model was fit to the retention test effect size data with a one-tailed *p*-value cutpoint of .025 included in the model (Vevea & Hedges, 1995). The adjusted model provided a significantly better fit to the data than the naïve random effects model. The model estimated the selection-adjusted benefit of self-controlled learning as $g = .11$, a dramatic departure from the naïve estimate of $g = .44$. Two additional bias correction techniques were conducted to assess the sensitivity of this result to changes in correction methodology. The PEESE method estimated the effect at $g = .05$, while *p*-curve estimated $g = .04$, and neither analysis was able to rule out the null hypothesis.

In parallel to the meta-analysis described above, a *p*-curve was conducted on the reported significant results. The p-curve used somewhat different inclusion criteria focusing only on published, statistically significant results suggesting a self-controlled learning benefit. In addition, the p-curve included results reported for any dependent measure in an article, even if the focal measure (of this meta-analysis) was reported as non-significant. Therefore, the p-curve was more inclusive of evidence reported by authors as favouring a self-controlled benefit while ignoring experiments with null effects. The results revealed both significant right skew below $p = .025$ (two-tailed) and a p-curve that was significantly flatter than a distribution with an expected power of 33%. The evidence of right skew, indicating superiority of self-control relative to yoked conditions, was tenuous and did not survive the deletion of the most extreme result–an experiment that reported a benefit from self-control of $g = 2.16$ (Gabriele Wulf & Adams, 2014). The overall p-curve produced an estimate that the true power of the included experiments was 5%, leading to a rejection of the hypothesis that the experiments contained evidential value.

It appears from these analyses that the substantial self-controlled learning literature is, as of now, insufficient to provide evidence that self-controlled practice is more effective than a yoked practice. The bias correction techniques applied in this analysis are sensitive to unknown conditions, such as the true average effect size and the amount of true heterogeneity; although efforts were taken to provide coverage across most plausible conditions. The corrected estimates produced by the weight-function model, p-curve, and PEESE methods appeared to converge on trivially small effects. Further, the p-curve of significant results suggested a lack of evidential value. Based on the model performance parameters we tested (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019), our results are consistent with a self-controlled learning benefit ranging from g = -.11 to .26, with a plausible upper 95% confidence limit of $g = .33$. Thus, this analysis does not rule out the possibility that self-controlled practice provides meaningful motor learning benefits on average. The present literature, however, appears insufficient to establish that a self-control benefit indeed exists.

Turning to the current theoretical debates surrounding the motivational and informational underpinnings of self-controlled learning, these debates now seem moot, or at least premature. The effectiveness of self-control was not moderated by choice-type, suggesting that self-controlled practice may be ineffective regardless of the nature of the choices provided. Indeed, the only factor we tested that moderated the effect of self-controlled practice was publication status.

Future studies

Given that the current meta-analysis failed to support the widely touted assertion of a substantial self-controlled learning benefit (Sanli, Patterson, Bray, & Lee, 2013; Ste-Marie, Carter, & Yantha, 2019; Gabriele Wulf & Lewthwaite, 2016), considerations need to be given to the design and research practices for future studies. Registered reports provide one possible path forward (Caldwell et al., 2020). A registered report involves submitting a research proposal to a two-phase peer-review. The first phase of the review occurs prior to data-collection and is assessed based on the proposed methodology, rationale, and potential contribution. If accepted in principle, researchers commit to carrying out the registered experiment and submitting the results in a final article for the second phase of peer-review. The final article is peer-reviewed for quality and adherence to the registered plan, but accept-reject decisions at this point are not based on the results. In theory, this practice should eliminate p-hacking and, for literatures composed entirely of registered reports, publication bias. A number of motor behaviour and/or kinesiology journals have begun adopting registered reports as an option for authors, including the *Human Movement Science*, *Frontiers in Movement Science and Sport Psychology*, *Journal of Sport and Exercise Psychology*, *Journal of Sport Sciences*, and *Registered Reports in Kinesiology*.

While registered reports are a potentially fruitful process to begin the accumulation of evidence regarding self-controlled learning, there are practical issues with investigating self-controlled learning that motor learning researchers may find overly burdensome. For example, to have 80% power to detect an effect of $g = .26$ with a two cell experimental design, 506 participants are required. If the weight-function adjusted estimate of $g = .11$ is accurate, $N = 2600$ are required. More challenging still would be testing between hypothesized motivational and informational mechanisms. For example, if a 2 (choice) X 2 (choice-relevance) experiment were conducted to test whether the instructional-relevance of choice fully attenuates its effect, four times as many participants would be required to maintain the same degree of power (Simonsohn, 2015). In contrast, the median sample size among experiments included in this meta-analysis was $N = 36$, which is typical of motor learning experiments in general (Lohse, Buchanan, & Miller, 2016).

In addition to challenges with establishing that an effect exists, additional challenges will emerge if researchers are interested in generalizing the benefits of self-controlled practice beyond comparisons to a yoked group, as has been the case thus far (Ste-Marie, Carter, & Yantha, 2019; Gabriele Wulf & Lewthwaite, 2016). Yoking may allow for inferences to be made about the act of making certain choices, but it may not provide an adequate control group for evaluating best practices in an applied setting (e.g., J. A. C. Barros, Yantha, Carter, Hussien, & Ste-Marie, 2019; Ste-Marie, Carter, & Yantha, 2019; Yantha, 2019). Indeed, given that our estimate suggests the advantage of self-controlled over yoked practice is small, if it exists at all, it seems unlikely that self-control would be more effective than an instructor-guided practice. An instructor-guided group could easily be argued to have advantages over a yoked group, because of the ability for the instructor to adapt choices to the current practice context and to make use of personal experience and expertise. Following this logic, experiments investigating the benefit of self-controlled over instructor-guided practice could conceivably require substantially larger samples than experiments that use yoked comparison groups.

Exploratory analysis of pre-registered experiments

There have been, to our knowledge, four pre-registered experiments that have compared self-controlled and yoked practice (Grand, Daou, Lohse, & Miller, 2017; McKay & Ste-Marie, 2020; St. Germain et al., 2021; Yantha, 2019). Three of these experiments failed to meet our inclusion criteria because they were not published or part of an accepted thesis at the time of the analysis (McKay & Ste-Marie, 2020; St. Germain et al., 2021; Yantha, 2019). These pre-registered experiments should provide estimates of the self-control effect unbiased by selection effects and are therefore more useful for estimating the real average effect than attempting to correct biased experiments after the fact (E. C. Carter, Schönbrodt, Gervais, & Hilgard, 2019). A random effects model was used to estimate the average effect of self-control in the four experiments and yielded $g = .02$, 95% CI [-.17, .21]. These results converge with the bias-corrected estimates around trivially small differences between self-controlled and yoked practice conditions.

Conclusion

We set out to assess the effect of self-controlled practice on motor learning. The published literature on the subject to date appeared unambiguously supportive of a self-control benefit, yet the results of this meta-analysis suggest this may not be the case. If authors, reviewers, and editors select for statistical significance when deciding if experiments get published, the published literature becomes biased (Ioannidis, 2005). Worse still, filtering based on statistical significance may well incentivize researchers to leverage researcher degrees of freedom to achieve a significant result, a practice known as p-hacking, further biasing the literature (Wicherts et al., 2016). An instructive example of the potential impact of selection effects comes from research studying the so-called ego-depletion effect (Roy F. Baumeister, Vohs, & Tice, 2007; Hagger, Wood, Stiff, & Chatzisarantis, 2010). In a typical study, participants are asked to engage in activities that supposedly drain a limited reservoir of willpower, termed ego-depletion, and are subsequently measured on a dependent measure requiring an additional exertion of self-control, such as a Stroop task. The typical finding is that performance suffers on the second task if ego-depletion occurs beforehand. A meta-analysis by Hagger and colleagues (2010) reported the average effect of ego-depleting interventions on willpower dependent measures was $d = .62$. There was apparent consensus in the field that willpower relied on a limited resource due to the ostensibly unambiguous evidence in support of the theory (R. F. Baumeister & Vohs, 2016). Nevertheless, when bias correction methods were applied in a meta-analysis of ego-depletion literature, the adjusted estimates often did not differ significantly from zero (E. C. Carter, Kofler, Forster, & McCullough, 2015). Subsequently, a pre-registered, multi-lab replication project tested a sample of $N = 2141$ and reported that the ego-depletion effect was close to zero (Hagger et al., 2016). Thus, a prominent psychological construct substantiated by a large corpus of peer-reviewed evidence was investigated using cutting edge meta-analytic techniques that corrected for selection bias and the result was a trivially small estimated effect–an estimate supported by a subsequent large scale pre-registered replication effort. Notably, both the bias corrected meta-analysis and the subsequent multi-lab replication efforts have been criticized by ego-depletion theorists (R. F. Baumeister & Vohs, 2016; Cunningham & Baumeister, 2016). Others have sharply challenged these critiques (Schimmack, 2020), and while debate continues among social psychologists about the underlying theory at stake (e.g., Dang, 2018), there is consensus that several methods shown to produce positive results in the past are unlikely to replicate in future experiments.

In stark parallel to the ego-depletion literature, the findings of the current research suggest the self-controlled motor learning literature may be similarly biased. As motor learning researchers consider the path forward for self-controlled learning, non-bias related limitations of the extant literature should be addressed. For example, yoked groups fail to isolate putative motivational and informational processes when self-controlling learners make choices pertinent to acquiring a skill (M. J. Carter, Rathwell, & Ste-Marie, 2016; M. J. Carter & Ste-Marie, 2017b; Lewthwaite, Chiviacowsky, Drews, & Wulf, 2015). Further, exclusive reliance on yoked comparison groups limits the generalizability of self-controlled learning to applied settings where the alternative to self-control is typically coach or instructor control (i.e., those with domain-specific knowledge). As motor learning researchers in this area move forward, they are faced with the question of whether this effect is worth the resources required to study it. If that answer is yes, then in addition to being pre-registered and an adequately powered design, future self-controlled learning experiments should provide insight about either the underlying processes at work or the real world usefulness of this practice variable.

Data, materials, and code availability

All material, data, and scripts to reproduce our analyses and figures can be accessed here: https://osf.io/qbg69.

Conflict of interest

The authors declare no competing interests.

Funding

MJC was supported by a Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant (RGPIN-2018-05589).

References

References marked with an asterisk (*) indicate studies included in the meta-analysis.

Abdollahipour, R., Palomo Nieto, M., Psotta, R., & Wulf, G. (2017). External focus of attention and autonomy support have additive benefits for motor performance in children. *Psychol. Sport Exerc.*, *32*, 17–24.

* Aiken, C. A., Fairbrother, J. T., & Post, P. G. (2012). The effects of self-controlled video feedback on the learning of the basketball set shot. *Front. Psychol.*, *3*, 338.

* Alami, A. (2013). *An examination of feedback request strategies when learning a multi-dimensional motor task under self-controlled and yoked conditions* (PhD thesis). University of Tennessee, Knoxville.

* Ali, A., Fawver, B., Kim, J., Fairbrother, J., & Janelle, C. M. (2012). Too much of a good thing: Random practice scheduling and self-control of feedback lead to unique but not additive learning benefits. *Front. Psychol.*, *3*, 503.

* Andrieux, M., Boutin, A., & Thon, B. (2016). Self-Control of task difficulty during early practice promotes motor skill learning. *J. Mot. Behav.*, *48*(1), 57–65.

* Andrieux, M., Danna, J., & Thon, B. (2012). Self-control of task difficulty during training enhances motor learning of a complex coincidence-anticipation task. *Res. Q. Exerc. Sport*, *83*(1), 27–35.

* Arsal, G. (2004). *Effects of external and Self-Controlled feedback schedule on retention of anticipation timing and ball throwing task* (Master's thesis). Middle East Technical University.

* Barros, J. A. (2010). *The effects of practice schedule and Self-Controlled feedback manipulations on the acquisition and retention of motor skills* (PhD thesis). University of Tennessee, Knoxville.

* Barros, J. A. C., Yantha, Z. D., Carter, M. J., Hussien, J., & Ste-Marie, D. M. (2019). Examining the impact of error estimation on the effects of self-controlled feedback. *Hum. Mov. Sci.*, *63*, 182–198.

Bartoš, F., & Schimmack, U. (2020). *Z-curve.2.0: Estimating replication rates and discovery rates.* PsyArXiv.

* Bass, A. D. (2015). *An experiment to chronometrically examine the effects of Self-Controlled feedback on the performance and learning of a sequential timing task* (Master's thesis). University of Tennessee, Knoxville.

* Bass, A. D. (2018). *The effect of observation on motor learning in a self-controlled feedback protocol* (PhD thesis). University of Tennessee, Knoxville.

Baumeister, R. F., & Vohs, K. D. (2016). Strength model of Self-Regulation as limited resource: Assessment, controversies, update (chapter 2). In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 54, pp. 67–127). Academic Press.

Baumeister, Roy F., Vohs, K. D., & Tice, D. M. (2007). The strength model of Self-Control. *Curr. Dir. Psychol. Sci.*, *16*(6), 351–355.

* Brydges, R., Carnahan, H., Safir, O., & Dubrowski, A. (2009). How effective is self-guided learning of clinical technical skills? It's all about process. *Med. Educ.*, *43*(6), 507–515.

* Bund, A., & Wiemeyer, J. (2004). Self-controlled learning of a complex motor skill: Effects of the learner's preferences on performance and self-efficacy. *Journal of Human Movement Studies*, *47*, 215–236.

Caldwell, A. R., Vigotsky, A. D., Tenan, M. S., Radel, R., Mellor, D. T., Kreutzer, A., … Consortium for Transparency in Exercise Science (COTES) Collaborators. (2020). Moving sport and exercise science forward: A call for the adoption of more transparent research practices. *Sports Med.*, *50*(3), 449–459.

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *J. Exp. Psychol. Gen.*, *144*(4), 796–815.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of Meta-Analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144.

Carter, M. J., Carlsen, A. N., & Ste-Marie, D. M. (2014). Self-controlled feedback is effective if it is based on the learner's performance: A replication and extension of chiviacowsky and wulf (2005). *Front. Psychol.*, *5*, 1325.

* Carter, M. J., & Patterson, J. T. (2012). Self-controlled knowledge of results: Age-related differences in motor learning, strategies, and error detection. *Hum. Mov. Sci.*, *31*(6), 1459–1472.

Carter, M. J., Rathwell, S., & Ste-Marie, D. (2016). Motor skill retention is modulated by strategy choice during self-controlled knowledge of results schedules. *Journal of Motor Learning and Development*, *4*(1), 100–115.

Carter, M. J., & Ste-Marie, D. M. (2017a). An interpolated activity during the knowledge-of-results delay interval eliminates the learning advantages of self-controlled feedback schedules. *Psychol. Res.*, *81*(2), 399–406.

Carter, M. J., & Ste-Marie, D. M. (2017b). Not all choices are created equal: Task-relevant choices enhance motor learning compared to task-irrelevant choices. *Psychon. Bull. Rev.*, *24*(6), 1879–1888.

* Chen, D. D., Hendrick, J. L., & Lidor, R. (2002). Enhancing self-controlled learning environments: The use of self-regulated feedback information. *Journal of Human Movement Studies*, *43*(1), 69.

* Chiviacowsky, S. (2014). Self-controlled practice: Autonomy protects perceptions of competence and enhances motor learning. *Psychol. Sport Exerc.*, *15*(5), 505–510.

* Chiviacowsky, S., & Lessa, H. T. (2017). Choices over feedback enhance motor learning in older adults. *Journal of Motor Learning and Development*, *5*(2), 304–318.

* Chiviacowsky, S., & Wulf, G. (2002). Self-controlled feedback: Does it enhance learning because performers get feedback when they need it? *Res. Q. Exerc. Sport*, *73*(4), 408–415.

Chiviacowsky, S., & Wulf, G. (2005). Self-controlled feedback is effective if it is based on the learner's performance. *Res. Q. Exerc. Sport*, *76*(1), 42–48.

Chiviacowsky, S., Wulf, G., & Lewthwaite, R. (2012). Self-controlled learning: The importance of protecting perceptions of competence. *Front. Psychol.*, *3*, 458.

* Chiviacowsky, S., Wulf, G., Lewthwaite, R., & Campos, T. (2012). Motor learning benefits of self-controlled practice in persons with parkinson's disease. *Gait Posture*, *35*(4), 601–605.

* Chiviacowsky, S., Wulf, G., Machado, C., & Rydberg, N. (2012). Self-controlled feedback enhances learning in adults with down syndrome. *Rev. Bras. Fisioter.*, *16*(3), 191–196.

* Chiviacowsky, S., Wulf, G., Medeiros, F. L. de, Kaefer, A., & Tani, G. (2008). Learning benefits of self-controlled knowledge of results in 10-year-old children. *Res. Q. Exerc. Sport*, *79*(3), 405–410.

Chua, L.-K., Wulf, G., & Lewthwaite, R. (2018). Onward and upward: Optimizing motor performance. *Hum. Mov. Sci.*, *60*, 107–114.

Coburn, K. M., & Vevea, J. L. (2017). Weightr: Estimating weight-function models for publication bias. *R Package Version*, *1*(2).

Couvillion, K. F., Bass, A. D., & Fairbrother, J. T. (2020). Increased cognitive load during acquisition of a continuous task eliminates the learning effects of self-controlled knowledge of results. *J. Sports Sci.*, *38*(1), 94–99.

Cunningham, M. R., & Baumeister, R. F. (2016). How to make nothing out of something: Analyses of the impact of study sampling and statistical interpretation in misleading Meta-Analytic conclusions. *Front. Psychol.*, *7*, 1639.

Dang, J. (2018). An updated meta-analysis of the ego depletion effect. *Psychol. Res.*, *82*(4), 645–651.

* Davis, J. (2009). *Effects of self-controlled feedback on the squat* (Master's thesis). State University of New York College at Cortland.

* Fagundes, J., Chen, D. D., & Laguna, P. (2013). Self-control and frequency of model presentation: Effects on learning a ballet passé relevé. *Hum. Mov. Sci.*, *32*(4), 847–856.

* Fairbrother, J. T., Laughlin, D. D., & Nguyen, T. V. (2012). Self-controlled feedback facilitates motor learning in both high and low activity individuals. *Front. Psychol.*, *3*, 323.

* Ferreira, B. P., Malloy-Diniz, L. F., Parma, J. O., Nogueira, N. G. H. M., Apolinário-Souza, T., Ugrinowitsch, H., & Lage, G. M. (2019). Self-Controlled feedback and learner impulsivity in sequential motor learning. *Percept. Mot. Skills*, *126*(1), 157–179.

* Figueiredo, L. S., Ugrinowitsch, H., Freire, A. B., Shea, J. B., & Benda, R. N. (2018). External control of knowledge of results: Learner involvement enhances motor skill transfer. *Percept. Mot. Skills*, *125*(2), 400–416.

* Ghorbani, S. (2019). Motivational effects of enhancing expectancies and autonomy for motor learning: An examination of the OPTIMAL theory. *J. Gen. Psychol.*, *146*(1), 79–92.

* Grand, K. F., Bruzi, A. T., Dyke, F. B., Godwin, M. M., Leiker, A. M., Thompson, A. G., … Miller, M. W. (2015). Why self-controlled feedback enhances motor learning: Answers from electroencephalography and indices of motivation. *Hum. Mov. Sci.*, *43*, 23–32.

* Grand, K. F., Daou, M., Lohse, K. R., & Miller, M. W. (2017). Investigating the mechanisms underlying the effects of an incidental choice on motor learning. *Journal of Motor Learning and Development*, *5*(2), 207–226.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A multilab preregistered replication of the Ego-Depletion effect. *Perspect. Psychol. Sci.*, *11*(4), 546–573.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychol. Bull.*, *136*(4), 495–525.

Halperin, I., Chapman, D. W., Martin, D. T., Lewthwaite, R., & Wulf, G. (2017). Choices enhance punching performance of competitive kickboxers. *Psychol. Res.*, *81*(5), 1051–1058.

Halperin, I., Wulf, G., Vigotsky, A. D., Schoenfeld, B. J., & Behm, D. G. (2018). Autonomy: A missing ingredient of a successful program? *Strength & Conditioning Journal*, *40*(4), 18.

* Hansen, S., Pfeiffer, J., & Patterson, J. T. (2011). Self-control of feedback during motor learning: Accounting for the absolute amount of feedback using a yoked group with self-control over feedback. *J. Mot. Behav.*, *43*(2), 113–119.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). Doing meta-analysis in r: A hand-on guide. *PROTECT Lab Erlangen.*

* Hartman, J. M. (2007). Self-controlled use of a perceived physical assistance device during a balancing task. *Percept. Mot. Skills*, *104* (3 Pt 1), 1005–1016.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *J. Educ. Behav. Stat.*, *21* (4), 299–332.

* Hemayattalab, R., Arabameri, E., Pourazar, M., Ardakani, M. D., & Kashefi, M. (2013). Effects of self-controlled feedback on learning of a throwing task in children with spastic hemiplegic cerebral palsy. *Res. Dev. Disabil.*, *34* (9), 2884–2889.

Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons.

* Ho, R. L. M. (2016). *Self-controlled learning and differential goals: Does "too easy" and "too difficult" affect the self-control paradigm?* (PhD thesis). California State University, Long Beach.

* Holmberg, B. A. (2013). *The "when" and the "what": Effects of Self-Control of feedback about multiple critical movement features on motor performance and learning* (PhD thesis). University of Tennessee, Knoxville.

Hong, S., & Reed, W. R. (2021). Using monte carlo experiments to select meta-analytic estimators. *Research Synthesis Methods*, *12* (2), 192–215.

* Huet, M., Camachon, C., Fernandez, L., Jacobs, D. M., & Montagne, G. (2009). Self-controlled concurrent feedback and the education of attention towards perceptual invariants. *Hum. Mov. Sci.*, *28* (4), 450–467.

* Ikudome, S., Kou, K., Ogasa, K., Mori, S., & Nakamoto, H. (2019). The effect of choice on motor learning for learners with different levels of intrinsic motivation. *J. Sport Exerc.*

Psychol., *41*(3), 159–166.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.*, *2*(8), e124.

Iwatsuki, T., Navalta, J. W., & Wulf, G. (2019). Autonomy enhances running efficiency. *J. Sports Sci.*, *37*(6), 685–691.

* Jalalvand, M., Bahram, A., Daneshfar, A., & Arsham, S. (2019). The effect of gradual Self-Control of task difficulty and feedback on learning golf putting. *Res. Q. Exerc. Sport*, *90*(4), 429–439.

* Janelle, C. M., Barba, D. A., Frehlich, S. G., Tennant, L. K., & Cauraugh, J. H. (1997). Maximizing performance feedback effectiveness through videotape replay and a self-controlled learning environment. *Res. Q. Exerc. Sport*, *68*(4), 269–279.

Janelle, C. M., Kim, J., & Singer, R. N. (1995). Subject-controlled performance feedback and learning of a closed motor skill. *Percept. Mot. Skills*, *81*(2), 627–634.

* Jones, A. (2010). *Effects of amount and type of self-regulation opportunity during skill acquisition on motor learning* (PhD thesis). McMaster University.

Jowett, N., LeBlanc, V., Xeroulis, G., MacRae, H., & Dubrowski, A. (2007). Surgical skill acquisition with self-directed practice using computer-based video training. *Am. J. Surg.*, *193*(2), 237–242.

* Kaefer, A., Chiviacowsky, S., Meira, C. de M., Jr, & Tani, G. (2014). Self-controlled practice enhances motor learning in introverts and extroverts. *Res. Q. Exerc. Sport*, *85*(2), 226–233.

* Keetch, K. M., & Lee, T. D. (2007). The effect of self-regulated and experimenter-imposed practice schedules on motor learning for tasks of varying difficulty. *Res. Q. Exerc. Sport*,

78(5), 476–486.

* Kim, Y., Kim, J., Kim, H., Kwon, M., Lee, M., & Park, S. (2019). Neural mechanism underlying self-controlled feedback on motor skill learning. *Hum. Mov. Sci.*, *66*, 198–208.

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434.

* Leiker, A. M., Bruzi, A. T., Miller, M. W., Nelson, M., Wegman, R., & Lohse, K. R. (2016). The effects of autonomous difficulty selection on engagement, motivation, and learning in a motion-controlled video game task. *Hum. Mov. Sci.*, *49*, 326–335.

* Leiker, A. M., Pathania, A., Miller, M. W., & Lohse, K. R. (2019). Exploring the neurophysiological effects of Self-Controlled practice in motor skill learning. *Journal of Motor Learning and Development*, *7*(1), 13–34.

* Lemos, A., Wulf, G., Lewthwaite, R., & Chiviacowsky, S. (2017). Autonomy support enhances performance expectancies, positive affect, and motor learning. *Psychol. Sport Exerc.*, *31*, 28–34.

* Lessa, H. T., & Chiviacowsky, S. (2015). Self-controlled practice benefits motor learning in older adults. *Hum. Mov. Sci.*, *40*, 372–380.

* Lewthwaite, R., Chiviacowsky, S., Drews, R., & Wulf, G. (2015). Choose to move: The motivational impact of autonomy support on motor learning. *Psychon. Bull. Rev.*, *22*(5), 1383–1388.

* Lim, S., Ali, A., Kim, W., Kim, J., Choi, S., & Radlo, S. J. (2015). Influence of self-controlled feedback on learning a serial motor skill. *Percept. Mot. Skills*, *120*(2), 462–474.

* Lindern, A. D. von. (2017). *Self-control effect during a reduction of feedback availability* (PhD thesis). University of Tennessee, Knoxville.

Lohse, K., Buchanan, T., & Miller, M. (2016). Underpowered and overworked: Problems with data analysis in motor learning studies. *Journal of Motor Learning and Development*, *4*(1), 37–58.

* Marques, P. G., & Corrêa, U. C. (2016). The effect of learner's control of self-observation strategies on learning of front crawl. *Acta Psychol.*, *164*, 151–156.

* Marques, P. G., Thon, R. A., Espanhol, J., Tani, G., & Corrêa, U. C. (2017). The intermediate learner's choice of self-as-a-model strategies and the eight-session practice in learning of the front crawl swim. *Kinesiology*, *49*(1).

McKay, B., & Ste-Marie, D. M. (2020). Autonomy support and reduced feedback frequency have trivial effects on learning and performance of a golf putting task. *Hum. Mov. Sci.*, *71*, 102612.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in Meta-Analysis: An evaluation of selection methods and some cautionary notes. *Perspect. Psychol. Sci.*, *11*(5), 730–749.

* Norouzi, E., Hossini, F. S., & Aghdasi, M. T. (2016). Effect of self-control feedback on the learning of a throwing task with emphasis on decision-making process. *Open Science Journal of Psychology*, *2*(6), 32.

* Nunes, M. E. de S., Correa, U. C., Souza, M. G. T. X. de, Basso, L., Coelho, D. B., & Santos, S. (2019). No improvement on the learning of golf putting by older persons with Self-Controlled knowledge of performance. *J. Aging Phys. Act.*, *27*(3), 300–308.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

* Ostrowski, J. (2015). *The influence of shame on the frequency of self-controlled feedback and motor learning* (Master's thesis). Southern Illinois University Carbondale.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychol. Bull.*, *134*(2), 270–300.

* Patterson, J. T., & Carter, M. (2010). Learner regulated knowledge of results during the acquisition of multiple timing goals. *Hum. Mov. Sci.*, *29*(2), 214–227.

* Patterson, J. T., Carter, M. J., & Hansen, S. (2013). Self-controlled KR schedules: Does repetition order matter? *Hum. Mov. Sci.*, *32*(4), 567–579.

* Patterson, J. T., Carter, M., & Sanli, E. (2011). Decreasing the proportion of self-control trials during the acquisition period does not compromise the learning advantages in a self-controlled context. *Res. Q. Exerc. Sport*, *82*(4), 624–633.

* Patterson, J. T., & Lee, T. D. (2010). Self-regulated frequency of augmented information in skill learning. *Can. J. Exp. Psychol.*, *64*(1), 33–40.

* Post, P. G., Aiken, C. A., Laughlin, D. D., & Fairbrother, J. T. (2016). Self-control over combined video feedback and modeling facilitates motor learning. *Hum. Mov. Sci.*, *47*, 49–59.

* Post, P. G., Fairbrother, J. T., & Barros, J. A. C. (2011). Self-controlled amount of practice benefits learning of a motor skill. *Res. Q. Exerc. Sport*, *82*(3), 474–481.

* Post, P. G., Fairbrother, J. T., Barros, J. A. C., & Kulpa, J. D. (2014). Self-Controlled practice within a fixed time period facilitates the learning of a basketball set shot.

Journal of Motor Learning and Development, *2*(1), 9–15.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Re, A. C. D. (2013). Compute.es: Compute effect sizes. In *R Package*. Retrieved from https://cran.r-project.org/package=compute.es

* Rydberg, N. (2011). *The effect of self-controlled practice on forearm passing, motivation, and affect in women's volleyball players* (Master's thesis). University of Nevada, Las Vegas.

* Sanli, E. A., & Patterson, J. T. (2013). Learning effects of self-controlled practice scheduling for children and adults: Are the advantages different? *Percept. Mot. Skills*, *116*(3), 741–749.

Sanli, E. A., Patterson, J. T., Bray, S. R., & Lee, T. D. (2013). Understanding self-controlled motor learning protocols through the Self-Determination theory. *Front. Psychol.*, *3*, 611.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-Analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Rev. Educ. Res.*, *84*(3), 328–364.

Schimmack, U. (2020). A meta-psychological perspective on the decade of replication failures in social psychology. *Canadian Psychology*, *61*(4), 364–376.

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychon. Bull. Rev.*, *20*(1), 21–53.

Simonsohn, U. (2015). [17] no-way interactions. *The Winnower*. https://doi.org/10.15200/winn.142559.90552

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *J. Exp. Psychol. Gen.*, *143*(2), 534–547.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve and effect size: Correcting for publication bias using only significant results. *Perspect. Psychol. Sci.*, *9*(6), 666–681.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to ulrich and miller (2015). *J. Exp. Psychol. Gen.*, *144*(6), 1146–1152.

St. Germain, L., Williams, A., Poskus, A., Balbaa, N., Leshchyshen, O., Lohse, K. R., & Carter, M. J. (2021). *Increased perceptions of autonomy through choice fail to enhance motor skill retention*. PsyArXiv.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias: T. D. STANLEY AND h. DOUCOULIAGOS. *Res. Synth. Methods*, *5*(1), 60–78.

Ste-Marie, D. M., Carter, M. J., & Yantha, Z. D. (2019). Self-controlled learning: Current findings, theoretical perspectives, and future directions. In N. J. Hodges & A. M. Williams (Eds.), *Skill acquisition in sport: Research, theory and practice* (pp. 119–140). Routledge.

* Ste-Marie, D. M., Vertes, K. A., Law, B., & Rymal, A. M. (2013). Learner-Controlled Self-Observation is advantageous for motor skill acquisition. *Front. Psychol.*, *3*, 556.

* Tsai, M.-J., & Jwo, H. (2015). CONTROLLING ABSOLUTE FREQUENCY OF FEEDBACK IN a SELF-CONTROLLED SITUATION ENHANCES MOTOR

LEARNING. *Percept. Mot. Skills*, *121*(3), 746–758.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychol. Methods*, *10*(4), 428–443.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.*, *36*(3), 1–48.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., Aert, R. C. M. van, & Assen, M. A. L. M. van. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Front. Psychol.*, *7*, 1832.

* Williams, C. K., Tseung, V., & Carnahan, H. (2017). Self-Control of haptic assistance for motor learning: Influences of frequency and opinion of utility. *Front. Psychol.*, *8*, 2082.

Woodard, K. F., & Fairbrother, J. T. (2020). Cognitive loading during and after continuous task execution alters the effects of self-controlled knowledge of results. *Front. Psychol.*, *11*, 1046.

* Wu, Wilbur Fong Wah. (2007). *Self-control of learning multiple motor skills* (PhD thesis). Louisiana State University.

* Wu, Will F. W., & Magill, R. A. (2011). Allowing learners to choose: Self-controlled practice schedules for learning multiple movement patterns. *Res. Q. Exerc. Sport*, *82*(3), 449–457.

Wulf, Gabriele. (2007). Self-controlled practice enhances motor learning: Implications for physiotherapy. *Physiotherapy*, *93*(2), 96–101.

Wulf, Gabriele, & Adams, N. (2014). Small choices can enhance balance learning. *Hum. Mov. Sci.*, *38*, 235–240.

* Wulf, Gabriele, Chiviacowsky, S., & Cardozo, P. L. (2014). Additive benefits of autonomy support and enhanced expectancies for motor learning. *Hum. Mov. Sci.*, *37*, 12–20.

* Wulf, Gabriele, Chiviacowsky, S., & Drews, R. (2015). External focus and autonomy support: Two important factors in motor learning have additive benefits. *Human Movement Science*, *40*, 176–184.

* Wulf, G., Clauss, A., Shea, C. H., & Whitacre, C. A. (2001). Benefits of self-control in dyad practice. *Res. Q. Exerc. Sport*, *72*(3), 299–303.

* Wulf, Gabriele, Iwatsuki, T., Machin, B., Kellogg, J., Copeland, C., & Lewthwaite, R. (2018). Lassoing skill through learner choice. *J. Mot. Behav.*, *50*(3), 285–292.

Wulf, Gabriele, & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychon. Bull. Rev.*, *23*(5), 1382–1414.

Wulf, Gabriele, & Mornell, A. (2008). Insights about practice from the perspective of motor learning: A review. *Music Performance Research*, *2*, 1–25.

* Wulf, Gabriele, Raupach, M., & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Res. Q. Exerc. Sport*, *76*(1), 107–111.

Wulf, Gabriele, Shea, C., & Lewthwaite, R. (2010). Motor skill learning and performance: A review of influential factors. *Med. Educ.*, *44*(1), 75–84.

* Wulf, G., & Toole, T. (1999). Physical assistance devices in complex motor skill learning: Benefits of a self-controlled practice schedule. *Res. Q. Exerc. Sport*, *70*(3), 265–272.

Yantha, Z. (2019). *The recommendation for learners to be provided with control over their feedback schedule is questioned in a self-controlled learning paradigm* (Master's thesis). Université d'Ottawa / University of Ottawa.

Appendix A

P-curve disclosure form

Table A1

Experiment information from papers included in the p-curve analysis.

Original paper	Quoted text from original paper indicated predicted benefit of self-control relative to yoked practice	Design	Key statistical result	Quoted text from original paper with statistical results	Result
Andrieux, Danna & Thon (2012)	"Thus, we hypothesized that a practice condition in which the learner could set the level of task difficulty would be more beneficial for learning than a condition in which this parameter was imposed."	Two cell	Difference in means	"A follow up analysis restricted to the first two blocks revealed a significant difference between groups, $F(1, 36) = 4.85$, $p < .05$, partial eta squared = .12. Self-controlled learners were significantly more accurate (M AE = 12.73 mm, SE = 1.57) than their yoked counterparts (M AE = 18.1 mm, SE = 1.87) after a 24-hr rest."	$F(1, 36) = 4.85$
Andrieux, Boutin, & Thon (2016)	"Two main reasons led us to expect that self-control of nominal task difficulty would enhance motor skill learning, and especially when introduced during early practice rather than during late practice."	Four cell (Full self-control, full yoked, self-control then yoked, yoked then self-control)	Difference in means	"Planned pairwise comparisons revealed that the self-control groups exhibited lower RMSE (SC + SC, SC + YO, and YO + SC groups) than their yoked group counterparts (YO + YO group), $F(1, 44) = 14.02$, $p < .01$."	$F(1, 44) = 14.02$

Brydges, Carnahan, Safir & Dubrowski (2009)	"We hypothesised that participants with self-guided access to instruction would learn more than participants whose access to instruction was externally controlled."	2 (Control: self, yoked) X 2 (Goals: process, outcome)	Difference in means	"The self-process group performed better on the retention test than the control-process group (Fig. 1). This effect was significant for time taken, (F[1,23] = 4.33, P < 0.05)."	$F(1,23) = 4.33$
Chiviacowsky (2014)	"We hypothesized that participants of the self-controlled group would show superior motor learning than yoked participants"	Two cell	Difference in means	"The Self group outperformed the Yoked group. The group main effect was significant, $t(26) = 2.08$, $p = .04$, $d = .78$."	$t(26) = 2.08$
Chiviacowsky, Wulf, de Medeiros, Kaefer & Tani (2008)	"Therefore, the purpose of the present study was to examine whether the learning benefits of self-controlled KR would generalize to children."	Two cell	Difference in means	"The self-control group had higher accuracy scores than the yoked group. This difference was significant, $F(1, 24) = 4.40$, $p < .05$."	$F(1, 24) = 4.40$
Chiviacowsky, Wulf, Lewthwaite, & Campos (2012)	"The potential benefits of self-controlled practice have yet to be examined in persons with PD...under the assumption that self-controlled practice would enhance the learning of the task..."	Two cell	Difference in means	"The self-control group was overall more effective than the yoked group. Time in balance was significantly longer for the self-control group, $F(1, 26) = 4.25$, $p < .05$."	$F(1, 26) = 4.25$

Chiviacowsky Wulf, Machado & Rydberg (2012)	"We predicted that self-controlled practice, in particular the ability to choose when to receive feedback, would result in more effective learning compared to a practice condition without this opportunity (yoked group)."	Two cell	Difference in means	"The day following practice, a retention test (without feedback) revealed lower AEs for the self-control group than the yoked group (see Figure 2, right). The group difference was significant, with $F(1, 28)= 4.72$, $p < 0.05$, eta squared =.14."	$F(1, 28)= 4.72$
Hartman (2007)	"The primary aim of this study was to test whether there would exist a learning advantage for a self-controlled group, as opposed to a yoked control group, for learning a dynamic balance task."	Two cell	Difference in means	"To assess the relatively permanent or learning effects of practice with or without a self-controlled use of a balance pole, both groups performed a retention test on Day 3. The group effect was significant, $F(1, 17) = 8.29$, $p < .01$, with the Self-control group outperforming the yoked group."	$F(1, 17) = 8.29$
Kaefer, Chivia-cowsky, Meira Jr. & Tani (2014)	"...both self-controlled groups (introverts and extroverts) will achieve a level of activation that facilitates learning through the control of stimulation source (feedback) in comparison with the groups that do not have control over it."	2 (Control: self, yoked) X 2 (Personality: introvert, extrovert)	Difference in means	"The groups' main effects were detected on the factor "feedback type": Self-controlled groups performed better, $F(1, 52) = 4.13$, $p < .05$, compared with externally controlled groups"	$F(1, 52) = 4.13$

Leiker, Bruzi, Miller, Nelson, Wegman & Lohse (2016)	"We hypothesized that participants in the self-controlled group would show superior learning (i.e., better performance on retention and transfer tests) compared to the yoked group."	Two cell	Difference in means	"Controlling for pre-pest, there was a significant main effect of group, $F(1,57) = 4.51$, $p = 0.04$, partial eta squared = 0.07, such that participants in the self-controlled group performed better on the post-test than participants in the yoked group."	$F(1,57) = 4.51$
Lemos, Wulf, Lewthwaite & Chiviacowsky (2017)	"Independent of which factor the learner is given control over e or whether or not this factor is directly related to the task to be learned e the learning benefits appear to be very robust."	Two cell	Difference in means	"On the retention test, choice participants clearly outperformed the control group. The group main effect was significant, $F(1, 22) = 88.16$, $p < 0.01$."	$F(1, 22) = 88.16$
Lessa & Chiviacowsky (2015)	"...it was hypothesized that older adult participants of the self-group would demonstrate superior motor learning results, presenting faster task times on the speed cup-stacking task, when compared with participants in the yoked control group."	Two cell	Difference in means	"The analysis of the retention test revealed significant differences between groups, $F(1,34) = 4.87$, $p < .05$...with participants of the self-control group presenting faster task times compared to yoked participants."	$F(1,34) = 4.87$

Lewthwaite, Chiviacowsky, Drews & Wulf (2015; Exp. 1)	"In the present experiment, the choice learners were given was not related to task performance per se. Therefore, any learning benefits resulting from having, as opposed to not having, a choice would suggest that motivational factors are responsible for those effects."	Two cell	Difference in means	"On the retention test, during which white golf balls were used, the choice group showed significantly higher putting accuracy (36.8) than the yoked group (26.4), $F(1, 22) = 7.31$, $p < .05$"	$F(1, 22) = 7.31$
Lewthwaite, Chiviacowsky, Drews & Wulf (2015; Exp. 2)	"Given the potential theoretical importance of the finding in Experiment 1, we wanted to replicate it with another task and different type of choice."	Two cell	Difference in means	"On the retention test 1 day later, the choice group demonstrated significantly longer times in balance than the yoked group, $F(1, 27) = 7.93$, $p < .01$."	$F(1, 27) = 7.93$
Lim, Ali, Kim, Choi & Radlo (2015)	"It was expected that a self-controlled feedback schedule would be more effective for the learning and performance of serial skills for both acquisition and retention phases than a yoked schedule."	Two cell	Difference in means	"In the retention phase, there was a significant main effect for Group ($F(1, 22) = 18.27$, $p < .05$). The follow-up test indicated that the Self-controlled feedback group had higher performance (Cohen's $d = 6.4$) than the Yoked-feedback group during the retention test in both blocks."	$F(1, 22) = 18.27$

Patterson, Carter & Sanli (2011: Comparison 1)	"We expected that the structure of this self-controlled practice context would either add to or compromise the existing benefits attributed to a self-controlled practice context."	2 (Control: self, yoked) X 3 (Structure: full, all, faded)	Difference in means	"Specifically, the Self-Self condition demonstrated less |CE| compared to their Yoked-Yoked counterparts. This main effect was significant, $F(1, 18) = 8.06$, $p < .05$."	$F(1, 18) = 8.06$
Patterson, Carter & Sanli (2011: Comparison 2)	"We expected that the structure of this self-controlled practice context would either add to or compromise the existing benefits attributed to a self-controlled practice context."	2 (Control: self, yoked) X 3 (Structure: full, all, faded)	Difference in means	"The All-Self condition demonstrated less |CE| compared to the All-Yoked condition. This main effect was also statistically significant, $F(1, 18) = 4.67$, $p < .05$."	$F(1, 18) = 4.67$
Patterson, Carter & Sanli (2011: Comparison 3)	"We expected that the structure of this self-controlled practice context would either add to or compromise the existing benefits attributed to a self-controlled practice context."	2 (Control: self, yoked) X 3 (Structure: full, all, faded)	Difference in means	"The Faded-Self condition demonstrated less |CE| compared to the Faded-Yoked condition, supported by a main effect for group, $F(1, 18) = 5.78$, $p < .05$."	$F(1, 18) = 5.78$
Post, Fairbrother, Barros & Kulpa (2014)	"It was hypothesized that learners in the SC group would demonstrate superior accuracy and form scores compared with the yoked group during the retention test."	Two cell	Difference in means	"The univariate ANOVA for retention revealed a significant group effect, $F(1, 29) = 6.08$, $p = .020$. The SC group had higher Accuracy scores the YK group"	$F(1, 29) = 6.08$

Ste-Marie, Vertes, Law & Rymal (2013)	"We hypothesized that the Learner Controlled group would show superior physical performance of the trampoline skills... compared to the Experimenter Controlled group."	Two cell	Difference in means	"A separate independent samples t-test showed that the Learner Controlled group had significantly higher performance scores compared to the Experimenter Controlled group at retention, $t(58) = 3.21$, $p < .05$, $d = .753$."	$t(58) = 3.21$
Wulf & Adams (2014)	"We asked whether giving performers an incidental choice would also result in more effective learning of exercise routines."	2(Group: self-control, yoked) X 3 (Exercise: toe touch, head turn, ball pass) X 2 (Leg: left, right) mixed design with repeated measures on the final two factors	Difference in means	"On the retention test... the choice group showed fewer errors than the control group. The main effects of group, $F(1,18) = 25.35$, $p < .001$, was significant."	$F(1,18) = 25.35$
Wulf & Toole (1999)	"If the beneficial effects of self-control found in previous studies are more general in nature (i.e., some general mechanism responsible for these effects), learning advantage would also be expected for self-controlled use of physical assistance."	Two cell	Difference in means	"The main effect of Group, $F(1,24) = 4.54$, $p < .05$, was significant. Thus, allowing learners to select their own schedule of physical assistance during practice had a clearly beneficial effect on learning."	$F(1,24) = 4.54$

Wulf, Clauss, Shea & Whitacre (2001)	"Importantly, however, if self-control promotes the development of a more efficient movement technique, one should see greater movement efficiency, as indicated by delayed force onsets, in self-control as compared to yoked participants."	Two cell	Difference in means	"Whereas the self-control group demonstrated relative force onsets that, on average, occurred about half the distance between the center of the apparatus and the participant's maximum amplitude, the yoked group's average force onset had already occurred after they had travelled less than 20% of the distance to the maximum amplitude. This group difference was significant, $F(1,24) = 4.43$, $p < .05$."	$F(1,24) = 4.43$
Wulf, Raupach & Pfeiffer (2005)	"Thus, if the learning advantages of self-controlled practice generalize to observational practice, allowing learners to decide when they want to view a model presentation should result in enhanced retention performance, with regard to movement form and, perhaps, movement accuracy, compared to that of yoked learners."	Two cell	Difference in means	"Overall, the self-control group had higher form scores than the yoked group throughout retention. The main effect of group $F(1,23) = 5.16$, $p < .05$, was significant."	$F(1,23) = 5.16$

Wulf, Iwatsuki, Machin, Kellogg, Copeland, & Lewthwaite (2017) Exp 1.	"The purpose of the present experiments was threefold. First, we deemed it important to provide further evidence for the impact of incidental choices on motor skill learning. Given that self-controlled practice benefits for learning have frequently been interpreted from an information-processing perspective (e.g., Carter, Carlson, & Ste-Marie, 2014; Carter & Ste-Marie, 2016), with limited regard for rewarding-motivational explanations, further experimental evidence for learning enhancements through choices not directly related to the task seemed desirable (Experiments 1 and 2)."	Two cell	Difference in means	"On the retention test one day later, the choice group demonstrated higher scores than did the control group. The group effect was significant, $F(1, 29) = 5.72$, $p < .05$."	$F(1, 29) = 5.72$
Wulf, Chiviacowsky & Drews (2015)	"To summarize, we hypothesized that an external focus and autonomy support would have additive benefits for motor learning (i.e., retention and transfer performance), as evidenced by main effects for each factor."	2 (Autonomy support: self, yoked) X 2 (Focus: external, internal)	Difference in means	"On the retention test, the main effect of Autonomy Support was significant, $F(1, 64) = 6.98$, $p < .01$."	$F(1,64) = 6.98$

Ikudome, Kuo, Ogasa, Mori & Nakamoto (2019; Exp. 2)	"Previous studies manipulating participants' choice of variables relevant to the experimental task have indicated that such choices have a positive effect on motor learning due to deeper information processing by the participants. Based on these studies, it is possible that this positive effect would be observed regardless of participants' levels of intrinsic motivation, because this type of choice would not induce a change in perceived locus of causality from internal to external."	2(Choice: self, yoked) X 2 (Motivation: high, low)	Difference in means	"An ANCOVA indicated significant main effects of choice, $F(1, 39) = 8.93$, $p = .005$."	$F(1,39) = 8.93$

Note. KR = Knowledge of results; PD = Parkinson's disease; SC = Self-controlled

Appendix B

Missing data

Of the 78 experiments that met the eligibility criteria of this meta-analysis, 25 were excluded because of missing data. Those 25 experiments included 13 experiments that reported a statistically significant result, along with 12 that failed to find a significant self-controlled learning effect. Among the 13 experiments with missing data reporting a significant self-control benefit, one reported an inappropriate analysis (Hemayattalab et al.2013),[6] one reported statistics that do not match the experimental design (Jalalvand et al., 2019),[7] one reported significant effects on a partial analysis of their data rather than overall (Brydges et al., 2009), and one was previously identified by Lohse and colleagues (2016) as an outlier study (Carter & Patterson, 2012). The meta-analysis may have been strengthened by the exclusion of these results (Stanley et al., 2010).

Among the remaining nine experiments reporting a significant effect with missing data, two reported effects collapsed across immediate and delayed retention only (Patterson et al., 2013; Wu & Magill, 2011), two reported null effects on a higher priority measure and did not include sufficient data to calculate the effect size, while reporting a significant effect on a lower priority measur((Wulf et al., 2001; Wulf et al., 2005, both studies were included in the primary *p*-curve analysis), and five compared three or more groups in an omnibus ANOVA and reported the group effect as significant but did not include sufficient data to calculate the effect size for the self-control versus yoked comparison (Chen et al., 2002; Ghorbani, 2019; Huet et al., 2009; Janelle et al., 1997; Norouzi et al., 2016)

---

[6] Although data were collected in one dimension using concentric circles, AE and a measure of dispersion were analyzed together in a MANOVA. This measure of dispersion is not an accurate reflection of variability on a two-dimensional task for reasons described by Hancock et al. (1995).

[7] A subgroup analysis involving two groups $n = 15$ was reported with $df = 56$. The article reports r^2 effect sizes associated with each test that cannot be reproduced with the reported statistics or best guesses.