The combination of reporting bias and underpowered study designs has substantially exaggerated the motor learning benefits of self-controlled practice and enhanced expectancies: A meta-analysis

Abstract

Enhanced expectancies and autonomy-support through self-controlled practice conditions form the motivation pillar of OPTIMAL theory (Wulf & Lewthwaite, 2016). The influence of these practice variables on motor learning was recently evaluated in two separate meta-analyses. Both meta-analyses found that the published literature suggested a moderate and significant benefit on motor learning; however, evidence for reporting bias was found in both literatures. Although multiple bias-corrected estimates were reported in the self-controlled meta-analysis, there was no principled way to prefer one over the other. In the enhanced expectancies meta-analysis, the trim-and-fill-technique failed to correct the estimated effects. Here, we addressed these limitations by reanalyzing the data from both meta-analyses using robust Bayesian meta-analysis methods. Our reanalysis revealed that reporting bias substantially exaggerated the benefits of these practice variables in the original meta-analyses. The true effects appear small, uncertain, and potentially null. We found the estimated average statistical power among all studies from the original meta-analyses was 6% (95% confidence interval [5%, 13%]). These results provide compelling and converging evidence that strongly suggests the available literature is insufficient to support the motivation pillar of OPTIMAL theory. Our results highlight the need for adequately powered experimental designs if motor learning scientists want to make evidence-based recommendations.

Keywords: OPTIMAL theory, Motivation, RoBMA, *Z*-curve, Publication bias

Word count: 6354

For nearly 20 years, motor learning and performance research has been accumulating that some (e.g., Wulf et al., 2010; Wulf & Lewthwaite, 2016, 2021) have argued is not well-explained by classic information-processing-based theories (Guadagnoli & Lee, 2004; Salmoni et al., 1984; Schmidt, 1975). Examples range from putative motor learning benefits from receiving bogus social comparison feedback (e.g., Ávila et al., 2012) to ostensive learning and performance enhancement as a consequence of making incidental choices during practice (e.g., Lewthwaite et al., 2015). In the Optimizing Performance Through Intrinsic Motivation and Attention for Learning (OPTIMAL) theory, Wulf and Lewthwaite (2016) argued that information processing theories account for only a temporary influence of motivation, but fail to capture the more permanent influences of motivation on motor learning suggested by these findings; this failure to account for more permanent motor learning effects created the need for a new theory. The OPTIMAL theory provided an account of these, and numerous other lines of research, proposing that autonomy support and enhanced expectancies increase motivation, which explains the learning and performance benefits observed with self-controlled practice (see Sanli et al., 2013 for a review), positive feedback (e.g., Chiviacowsky & Wulf, 2007), social comparative feedback (e.g., Ávila et al., 2012), perceived task difficulty (e.g., C. Lee et al., 2011), conceptions of ability (e.g., Cimpian et al., 2007), self-modeling (e.g., Ste-Marie et al., 2011), and extrinsic rewards (e.g., Abe et al., 2011). While motivation had been deemphasized in previous motor learning theories, within OPTIMAL theory numerous lines of 21st century research may also be explained by motivational mechanisms, rather than or in addition to informational mechanisms.

Unfortunately, it has become recognized throughout social science that reporting bias in the literature can seriously distort the evidence (Collaboration, 2015; Fanelli, 2010; Hagger et al., 2016; Munafò et al., 2017). The results explained by the motivational factors in OPTIMAL theory may therefore be exaggerated or even non-existent (Gelman & Carlin, 2014). Research on the motor learning literature specifically has found evidence of the pernicious combination of low statistical power, high multiplicity (i.e., many statistical tests), and selective reporting (Lohse et al., 2016). Therefore, it is necessary to carefully examine the literature supporting OPTIMAL theory predictions to identify and potentially correct for selective reporting and publication.

The evidence underpinning the autonomy support predictions in OPTIMAL theory has been questioned in a recent meta-analysis, suggesting that low power combined with reporting bias may be responsible for the apparent benefits of self-controlled practice (McKay, Yantha, et al., 2022). In that study, it was observed that published experiments found an average self-controlled practice benefit of g = .54. However, several models of reporting bias provided a better fit to the data than the naive random effects model. Each of the models suggested the true average effect was small or potentially zero. Self-controlled practice was the primary literature addressed by the autonomy support factor in OPTIMAL theory; yet, these findings call into question whether there was even a robust phenomenon to explain.

Bacelar, Parma, Murrah, et al. (2022) investigated the literature addressed by the second motivation factor in OPTIMAL theory—enhanced expectancies. In their meta-analysis, it was found that the average benefit of studies that manipulated expectancies via interventions described in OPTIMAL theory was also g = .54. However, there was evidence of reporting bias that could not be accounted for with moderators in the study. The authors applied the trim-and-fill method in an effort to adjust for reporting bias, but it made no corrections and no other corrections were applied. It is noteworthy that self-controlled practice studies and the studies included in the enhanced expectancies meta-analysis have much in common (in addition to identical random effects point estimates based on the published literature): a) both literatures are comprised of variables predicted by OPTIMAL theory to increase motivation and in turn motor performance and learning; b) both include a study or two examined in a meta-analysis by Lohse and colleagues (2016) that found evidence of low power, multiplicity, and bias; and c) both showed signs of small study effects in their funnel plots. If both literatures have been affected similarly by reporting bias, then the current estimate of g = .54 for variables thought to enhance expectancies may be a substantial overestimate.

Addressing reporting bias presents substantial challenges to meta-analysts. Since reporting bias limits the information we have access to, it is impossible to know for certain how much bias is present or how large the impact is (Carter et al., 2019; McShane et al., 2016). The best we can do is think carefully about the mechanisms that potentially underlie reporting bias and attempt to model them accurately. In contrast, a naive random effects analysis assumes there is no reporting bias. Complicating matters, reporting bias can take several different forms with unique impacts on the final sample (Maier et al., 2022; Stefan & Schönbrodt, 2022; Thornton & Lee, 2000). To account for this, multiple models of reporting bias need to be attempted without knowing which is most likely *a priori*. Thus, bias-correction analyses are inherently sensitivity analyses (Mathur & VanderWeele, 2020; Sutton et al., 2000; Vevea & Woods, 2005). Until recently, if the results of multiple sensitivity analyses differed widely, there was no mechanism to reconcile the estimates. We now have Robust Bayesian Meta-Analysis (RoBMA) methods that apply Bayesian model averaging to allow meta-analysts to fit several plausible models (see Table 1 for descriptions) and give greater weight to the models that best account for the data (Bartoš et al., 2022; Maier et al., 2022). The RoBMA method provides single estimates of the average effect and heterogeneity, along with Bayes factors to quantify the evidence in support of a true effect, the presence of heterogeneity, and the presence of reporting bias. Results from simulation studies and analyses of real data with known reporting bias mechanisms suggest that RoBMA is substantially more accurate and less biased than naive random effects models, and also performs better than other competing reporting bias models (Bartoš et al., 2022).

[TABLE 1 NEAR HERE]

A challenge for most reporting bias models is large heterogeneity in true effects (Carter et al., 2019). While RoBMA appears to perform well with moderate levels of heterogeneity, its performance has not been evaluated when heterogeneity is high and the performance of each of its constituent reporting bias models suffers with high heterogeneity. The z-curve model was designed specifically to perform well regardless of heterogeneity (Bartoš & Schimmack, 2022). While z-curve does not provide adjusted effect size estimates (and such estimates may be meaningless with high heterogeneity), it instead estimates the average underlying power of included experiments. A significant difference between the estimated power of studies and the observed proportion of significant results can indicate the presence of reporting bias in a literature.

Here, we leveraged state-of-the-art robust Bayesian meta-analysis and *z*-curve methods to re-analyze the meta-analyses by McKay and colleagues (2022) and Bacelar and colleagues (2022). Considering the potential importance of OPTIMAL theory for the field of motor learning and performance (see T. D. Lee & Carnahan, 2021 for a discussion), it is imperative that the evidence buttressing its motivation predictions be evaluated as rigorously as possible. Critically, our re-analysis addresses limitations in both previous meta-analyses. First, the analysis of enhanced expectancies fit only one bias correction model—the trim-and-fill method (Duval & Tweedie, 1998)—and that model has been shown to result in exaggerated effect size estimates and severely inflated Type 1 error rates in the presence of publication bias and small or null effects (Bartoš et al., 2022; Carter et al., 2019; Hong & Reed, 2021). Second, although the results from multiple reporting bias models coalesced around small effect sizes (ranged from g = -.11 to g = .26) in the analysis of self-controlled practice, there are no principled reasons for preferring one estimate over another. We now consider a wider range of plausible models of reporting bias than that used in the previous meta-analyses. We also leverage Bayesian model averaging to upweight the best performing models, which has the advantage of evaluating single model-averaged posterior distributions for each parameter of interest. Lastly, we fit *z*-curve models to the data in both meta-analyses. With this technique, the average power can be estimated and compared to the rate of significant results, providing crucial insight into the quality of the evidence-base supporting motivation predictions in OPTIMAL theory.

Materials and methods

Data and code used in this study can be accessed using either of the following links: https://osf.io/vfza7/ or [redacted].

Description of datasets

The original meta-analyses followed similar inclusion criteria and data extraction policies (see Figure 1). All data were extracted by two independent researchers with conflicts settled by including a third researcher. The literature search for the self-controlled practice meta-analysis (for further details see McKay, Yantha, et al., 2022) ended in August 2019 and in June 2020 for the enhanced expectancies meta-analysis (for further details see Bacelar, Parma, Murrah, et al., 2022).

[FIGURE 1 NEAR HERE]

Robust Bayesian meta-analysis

The self-controlled practice and enhanced expectancy meta-data were reanalyzed using robust Bayesian meta-analysis with publication selection model-averaging (RoBMA-PSMA) (Bartoš et al., 2022). The RoBMA-PSMA method evaluates the evidence for reporting bias and adjusts effect size estimates and 95% credible intervals to account for the estimated bias. Since the true data generating process underlying the data is unknown, the RoBMA-PSMA method fits several plausible models to the data. The models vary with respect to whether a) an effect is present or absent, b) homogeneous studies are summarized with a common effect (fixed effects) or a distribution of heterogeneous studies summarized with a mean effect and heterogeneity (random effects), c) reporting bias is present or absent, and d) if present, how results are selected for publishing.

Two classes of models are included to account for selective publishing of results. The first class of models are known as selection models. In a selection model, a weight-function based on *p*-values is fit to the data and the probability that a result survives censorship to be included in the model is estimated for each *p*-value interval. For example, a one-tailed *p*-value cut point of .025 might be specified, corresponding to a two-sided *p*-value of .05 in the positive direction. The probability that non-significant results survive censorship is estimated relative to the probability that positive significant results are published, which is set at 1. Selection models can be used to model a variety of potential forms of selective publishing. For example, selection may be concerned with significance but not the direction of the effect; in which case a two-sided *p*-value of .05 can be fit to the model. Selection may also be based on both statistical significance in the positive direction and the direction of the point estimate; in which case one-sided *p*-values of .025 and .5 can be fit to the model. The RoBMA-PSMA method fits six different selection models corresponding to various plausible forms of selection based on discrete *p*-values.

The second class of models fit in the RoBMA-PSMA method model the relationship between observed effect sizes and their standard errors. A positive relationship between effect sizes and their standard errors has been termed ‘small study effects’ (Sterne et al., 2000). The RoBMA-PSMA includes the precision-effect test (PET) and the precision-effect test with standard errors (PEESE) for small study effects. The PET model fits the relationship between observed effect sizes and their standard errors, while the PEESE model instead includes the square of the standard errors (i.e., their variances) (Stanley & Doucouliagos, 2014). The difference between the PET and PEESE models is that the PET model fits a linear relationship between the effect observed in a study and the precision with which that effect was estimated. The PEESE model fits a quadratic relationship. Thus, the PET and PEESE models differ with respect to the assumed underlying selection process (Stanley & Doucouliagos, 2014). The PET model assumes that effect sizes increase in step with decreases in precision, consistent with selection based on statistical significance. The PEESE model assumes that studies with high precision are likely to be published regardless of statistical significance, whereas increasingly imprecise studies require increasingly larger effect size estimates to survive censorship.

RoBMA includes 36 separate models: a) six weight-function selection models in each level of a 2 (Effect: present, absent) x 2 (Heterogeneity: present, absent) matrix (24 models in total), b) PET and PEESE regression models in each level of the 2 x 2 Effect x Heterogeneity matrix (8 models in total), and c) models assuming no reporting bias at each level of the Effect x Heterogeneity matrix (4 models in total). The prior probabilities for the reporting-bias adjusted and unadjusted models are both set to .5; thus, summing to 1.0. The prior for the reporting-bias adjusted models is spread evenly between the selection and PET-PEESE regression model classes, so the priors for the two model classes sum to .5. Similarly, the prior distribution for the presence of an effect was divided evenly between a spike prior equal to zero effect (i.e., the null) and a normal distribution of plausible true effects with M = 0 and SD = 1. Likewise, the prior probabilities of a fixed effect and heterogenous effects were divided evenly, with plausible values of τ following an inverse gamma distribution with shape = 1 and scale = .15. We chose neutral priors[1] to allow the data to dominate our posterior estimates and not impose our own opinions on the analysis. Admittedly, our prior beliefs were that publication bias is likely, so the analysis we report is somewhat conservative in this respect relative to our personal priors. Estimates from all 36 models are combined using Bayesian model-averaging, which weights each model estimate based on its posterior probability. Models that better account for the data are given greater weight in the RoBMA-PSMA model and models that provide a poor fit are down-weighted. The adjusted effect size estimate is produced by averaging across all models. This preserves the uncertainty about the true data generating process while weighting the component models based on their relative performance.

In addition to providing an overall effect estimate, the RoBMA-PSMA model can also be used to calculate Bayes factors (BF) quantifying the strength of support for the presence (or absence) of an intervention effect, heterogeneity, or reporting bias. To illustrate, consider the question of whether there is or is not an intervention effect. Since we used a neutral prior that considered each possibility equally likely, we can take the ratio of the posterior probabilities of model ensembles that included an effect to those that did not. We always report the BFs so they can be interpreted as how many times more likely the data were assuming the best supported hypothesis compared to the other hypothesis. For example, a BF_{01} = 2.0 for the absence of an intervention effect suggests that the models assuming all variation is random fit the data twice as well as models

[1] We considered a more informed distribution of plausible effects based on the empirical benchmarks reported by Lovakov and Agadullina (2021) for social psychological phenomena. In this sensitivity analysis, we included a normal prior distribution with M = 0 and SD = .36. The results of this sensitivity analysis were nearly identical to the primary analysis.

predicting an effect is present.[2] Bayes factors should be interpreted as a continuous measure of the relative support for one hypothesis versus another, not as the probability a hypothesis is true.

RoBMA-PSMA outperforms other meta-analytic approaches, including each constituent model included in RoBMA-PSMA, in simulations covering a range of plausible scenarios (Maier et al., 2022). Further, an analysis of real data from an approximately known data generating process (multi-lab registered replication report) suggested that RoBMA-PSMA outperforms other available techniques (Bartoš et al., 2022). Simulations do suggest a simpler version of the RoBMA approach that includes only the PET-PEESE publication bias models (RoBMA-PP) is more effective in scenarios with strong *p*-hacking and publication bias. These strong performance indices combined with the ability to simultaneously model various plausible manifestations of reporting bias made RoBMA-PSMA (and RoBMA-PP) an attractive choice for re-analyzing the enhanced expectancies and self-controlled practice meta-analyses.

Z-curve

The self-controlled practice meta-data,[3] enhanced expectancy meta-data, and the combination of both the enhanced expectancy and self-controlled practice meta-data were analyzed with a *z*-curve. A *z*-curve analysis estimates the statistical power of all studies ever conducted within a given literature, even if those studies were not reported, on the basis of the significant results that are present (Bartoš & Schimmack, 2022). That power estimate is equivalent to the expected discovery rate, that is, the expected rate of significant results for a given literature. The expected discovery rate estimated by a *z*-curve and its corresponding 95% confidence interval can be compared to the observed discovery rate in the literature (the actual rate of significant results). A discrepancy between the 95% confidence interval of the expected discovery rate estimate and the observed

[2] BF_{10} is the inverse of BF_{01}, so while BF_{01} indicates evidence in support of the null hypothesis, BF_{10} indicates evidence in support of the alternative hypothesis.

[3] *Z*-curve results for self-controlled practice were reported previously in McKay, Yantha, et al. (2022). We reproduce them here for comparison to enhanced expectancies and motivational factors combined analyses.

discovery rate provides evidence of reporting bias.

Z-curve analyzes two-tailed *p*-values or absolute *z*-scores, which do not preserve the direction of the effect and therefore follow a folded normal distribution. Because the selection process that determines whether non-significant results survive censorship is unknown, *z*-curve includes only significant results. Therefore, the expected distribution of *z*-scores in a *z*-curve analysis is a folded normal distribution truncated at $z = 1.96$, corresponding to the conventional threshold for statistical significance. *Z*-curve is intended to be applied in both standard meta-analytic situations as well as broader investigations of entire fields, journals, or researcher publication histories. Therefore, the expected distribution of *z*-values in *z*-curve is heterogeneous, forming a mixture of truncated folded normal distributions with means equal to the population mean for each study and a standard deviation of 1. Critically, the mixture of truncated folded normal distributions for a given set of significant studies is a function of the average power of the population of studies from which they were sampled. By approximating this distribution *z*-curve can estimate the average power of all studies conducted within a given literature, the so-called expected discovery rate. *Z*-curve estimates the mixture model by using the expectation maximization algorithm (Dempster et al., 1977; G. Lee & Scott, 2012) to fit a finite mixture model of seven truncated folded normal distributions with population means of 0, 1, 2, 3, 4, 5, and 6.

Similar to RoBMA-PSMA (Maier et al., 2022), *z*-curve has also performed well in simulation studies and when applied to multi-lab registered replication data (Bartoš & Schimmack, 2022). *Z*-curve is a tool that can provide insight into the possible power shortcomings of a particular literature. Further, as *z*-curve was designed to accommodate highly heterogeneous inputs, it is ideal for exploring power and reporting bias among studies that used a variety of methods to manipulate the two motivational factors in Wulf and Lewthwaite's (2016) OPTIMAL theory of motor learning.

Data analysis

We fit two RoBMA-PSMA models to the enhanced expectancies meta-data using the effect sizes and variances calculated by Bacelar and colleagues (2022). The first model included all studies, while the second excluded two influential cases, consistent with the primary results reported in their original meta-analysis. The self-controlled practice effect sizes and standard errors reported by McKay and colleagues (2022) were analyzed using the same strategy.

Z-curve models were fit to the enhanced expectancy and self-controlled practice meta-data separately, as well as combined. The same strategy was followed regarding influential cases. For all analyses, the model excluding influential cases is reported in detail and models with all studies included are discussed only when there are meaningful differences.

Statistical analyses were conducted using R (Version 4.2.2; R Core Team, 2021) and the R-packages *geomtextpath* (Version 0.1.0; Cameron & van den Brand, 2022), *gt* (Version 0.6.0; Iannone et al., 2022), *invgamma* (Version 1.1; Kahle & Stamey, 2017), *metafor* (Version 3.4.0; Viechtbauer, 2010), *papaja* (Version 0.1.0.9999; Aust & Barth, 2020), *patchwork* (Version 1.1.0.9000; Pedersen, 2022), *plotly* (Version 4.10.0; Sievert, 2020), *pwr* (Version 1.3.0; Champely, 2020), *renv* (Version 0.15.5; Ushey, 2022), *RoBMA* (Version 2.2.2; Bartoš & Maier, 2020), *tidyverse* (Version 1.3.1; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), *truncdist* (Version 1.0.2; Novomestky & Nadarajah, 2016), and *zcurve* (Version 2.1.2; Bartoš & Schimmack, 2020) were used in this project.

Results

Robust Bayesian meta-analysis

Self-controlled practice

The model-averaged posterior distribution of the average effect from the RoBMA-PSMA model is displayed in Figure 2B. The results suggest moderate evidence against the presence of an effect,

BF_{01} = 3.16, very weak evidence against the presence of heterogeneity, BF_{rf} = 1.7, and overwhelming evidence for the presence of reporting bias, BF_{pb} = 18,399. The overall model ensemble estimated the effect of self-controlled practice as d = .034 (95% credible interval [.0, .248]). Heterogeneity was estimated as τ = .05 (95% credible interval [.0, .261]). A model fit with two influential cases[4] included found overwhelming evidence for the presence of heterogeneity, BF_{rf} = 1,924,516 and estimated τ = .559 (95% credible interval [.36, .78]). There were no other meaningful differences between models.

Enhanced expectancies

The model-averaged posterior distribution of the average effect from the RoBMA-PSMA model is displayed in Figure 2C. The results revealed weak evidence for the presence of an effect BF_{10} = 1.9, weak evidence for the presence of reporting bias, BF_{pb} = 2.3, and very strong evidence for the presence of heterogeneity, BF_{rf} = 47.6. The model ensemble estimated an average effect of d = .26 (95% credible interval [-.07, .63]). Heterogeneity was estimated as τ = .35 (95% credible interval [.07, .54]). A model fit with two influential cases[5] included found weak evidence for the absence of an effect BF_{01} = 2.5, strong evidence for the presence of reporting bias, BF_{pb} = 21, and overwhelming evidence for the presence of heterogeneity, BF_{rf} = 45,300. The model ensemble estimated an average effect of d = .00 (95% credible interval [-.62, .59]), τ = .49 (95% credible interval [.32, .68]).

[4] The two outliers in the self-controlled practice were studies by Lemos and colleagues (2017) and Marques and colleagues (2017). Lemos et al. 2017 measured ballet movement form and tested the effect of choosing among video demonstrations while Marques et al. 2017 measured front crawl movement form and tested the effect of choosing which self-modeling video to watch during practice.

[5] The two outliers in the enhanced expectancies meta-analysis were Goudini et al. (2018) and Navaee et al. (2018). The study by Goudini and colleagues measured linear tracing performance and tested the effect of feedback after good and bad trials. The study by Navaee and colleagues measured beanbag tossing performance and tested the effect of providing children on the autism spectrum with nonveridical feedback suggesting performance was 20% better than the average.

The RoBMA-PSMA models with and without influential cases differed primarily with respect to the fit of the PEESE models. When two large effect sizes with large standard errors were included in the analysis, the PEESE model provided a very strong fit to the data (BF_{10} = 39.9) when assuming no true effect but the presence of heterogeneity. When the two influential cases were excluded, the best fitting model was the PET under the same assumptions (BF_{10} = 9.4).[6]

[FIGURE 2 NEAR HERE]

Z-curve

Self-controlled practice

The results of the *z*-curve analysis can be seen in Figure 3B. The analysis estimated the average statistical power (expected discovery rate) of all experiments examining the effect of self-controlled practice was 6% (95% confidence intervals [5%, 12%]). Since the upper bound of the expected discovery rate does not overlap with the observed discovery rate of 48% (95% confidence interval [35%, 62%]), there is evidence of significant reporting bias. The estimated conditional power of the statistically significant results (expected replication rate) was 11% (95% confidence interval [3%, 30%]). Including influential cases did not markedly change the results.

Enhanced expectancies

The results of the *z*-curve analysis can be seen in Figure 3C. The analysis estimated that the expected discovery rate of studies conducted on enhanced expectancies was 8% (95% confidence interval [5%, 18%]). The observed discovery rate was 44% (95% confidence interval [31%, 59%]). Since the upper bound of the expected discovery rate does not overlap with the lower bound of the observed discovery rate, there is evidence of significant reporting bias. The expected replication

[6] We also fit RoBMA-PP models that perform better than RoBMA-PSMA models in the presence of strong *p*-hacking and publication bias. The RoBMA-PP models did not lead to substantively different conclusions for the analysis of self-controlled practice or enhanced expectancies.

rate of the statistically significant results was 33% (95% confidence interval [8%, 59%]). Including influential cases did not meaningfully change the results.

Motivational factors in OPTIMAL theory

The results of the *z*-curve analysis of enhanced expectancy and self-controlled practice meta-data combined can be seen in Figure 3D. The expected discovery rate of all studies conducted on the motivational factors in OPTIMAL theory is 6% (95% confidence interval [5%, 13%]). The observed discovery rate was 46% (95% confidence interval [37%, 56%]). The lower bound of the observed discovery rate does not overlap with the upper bound of the expected discovery rate, providing evidence of significant reporting bias in this literature. The expected replication rate of the statistically significant results was 21% (95% confidence interval [4%, 39%]). Including influential cases did not meaningfully change the results.

[FIGURE 3 NEAR HERE]

Discussion

A theory is formed based on observations of study results and/or real-world phenomena. It is evaluated by subsequent studies testing hypotheses derived from the theory. Thus, a theory's basis and support depend on the evidential value of the relevant studies. The OPTIMAL theory of motor learning (Wulf & Lewthwaite, 2016) was established through study results showing that enhancing learners' expectancies and control over practice conditions improves learning and further supported by studies testing these hypotheses (Wulf & Lewthwaite, 2021). Two recent meta-analyses were conducted to appraise the evidential value of studies testing whether self-controlled practice (McKay, Yantha, et al., 2022) and/or enhanced expectancies (Bacelar, Parma, Murrah, et al., 2022) improve motor learning. McKay and colleagues' (2022) meta-analysis found self-controlled practice benefited motor learning when using a naive random-effects model of published studies (g = 0.54), but little evidence for a benefit was observed when a suite of bias correction techniques were employed (g's ranged from -0.11 to 0.26). McKay et al. also used a z-curve analysis to estimate the statistical power of self-controlled practice studies and found them to be severely underpowered (power = 6%, 95% confidence interval [5%, 13%]). Bacelar and colleagues' (2022) meta-analysis found evidence that enhanced expectancies improve motor learning when utilizing a naive random-effects model (g = 0.54). Despite evidence of bias (i.e., funnel plot asymmetry), applying a trim-and-fill bias correction technique did not change the naive random-effects model estimate. Crucially, however, the trim-and-fill bias correction technique only slightly reduces bias and Type I error (Carter et al., 2019). Bacelar et al. did not estimate the statistical power of enhanced expectancies studies; although they did note that the included studies had small sample sizes.

Our objective in the current study was to conduct a holistic assessment of the motivation pillar of OPTIMAL theory. Specifically, we used RoBMA-PSMA—a state-of-the-art bias correction technique—to evaluate the effect of self-controlled practice and enhanced expectancies on motor learning (see Figure 2). Additionally, we used z-curve analyses to estimate the statistical power of the enhanced expectancies studies alone and then combined with the self-controlled practice studies to assess the motivation pillar of OPTIMAL theory (see Figure 3). Using RoBMA-PSMA to model the effect of self-controlled practice on motor learning, we found moderate evidence against the presence of an effect (BF_{01} = 3.16). The model ensemble estimated a small average effect of d = .034 with a 95% credible interval that reached zero [.0, .248]. We also found overwhelming evidence for the presence of reporting bias (BF_{pb} = 18,399). When modeling the effect of enhanced expectancies on motor learning with RoBMA-PSMA, we found weak evidence for the presence of an effect (BF_{10} = 1.9). The model ensemble estimated a small average effect of d = .26 with a 95% credible interval that included zero [-.07, .63]. Weak evidence for the presence of reporting bias (BF_{pb} = 2.3) was also found. There was very strong support for the presence of heterogeneity in the enhanced expectancies literature (BF_{rf} = 47.6). The likely presence of heterogeneity suggests there is not one true effect, so individual studies may have been testing interventions with real benefits. However, this also means individual studies may have been testing interventions with real detriments as well, and we cannot discern which studies fall into which category. The estimated τ

of .35 suggests the possibility of substantial variation in the effect of enhanced expectancies, highlighting the importance of cautious interpretation given the effect modifiers remain unknown.

Heterogeneity can be problematic for modeling reporting bias with the selection and regression models employed in RoBMA-PSMA (Carter et al., 2019). Fortunately, the *z*-curve was designed for heterogeneous samples and may therefore be a better method of evaluating reporting bias in the enhanced expectancies data. The *z*-curve analysis estimated the average statistical power of enhanced expectancies studies to be 8%. The 95% confidence interval [5%, 18%] did not overlap with the 95% confidence interval of the observed discovery rate [31%, 59%], providing evidence of significant reporting bias. Visual inspection of the distribution of *z*-values reveals a large concentration of barely significant results and a dearth of barely not significant results (Figure 3B). This pattern is consistent with selective reporting, and it is difficult to imagine an alternate process that would generate these results. To evaluate the potential presence of reporting bias across both motivation factors, as well as estimate the average power of studies that have investigated these phenomena, we applied a *z*-curve to studies from both meta-analyses. The *z*-curve analysis combining the enhanced expectancies and self-controlled practice studies estimated the average statistical power to be 6%. The 95% confidence interval [5%, 13%] did not overlap with the 95% confidence interval of the observed discovery rate [37%, 56%], indicating significant reporting bias.

Our results are mostly consistent with other recent meta-analyses and pre-registered experiments with large sample sizes that have raised concerns about the state of evidence for the motivation pillar in OPTIMAL theory. Concerning self-controlled practice, the RoBMA-PSMA estimate of d =.034 is within the range of estimates reported by McKay et al. (2022), which showed g's ranging from -0.11 to 0.26. Our finding of overwhelming evidence for reporting bias is also consistent with McKay and colleagues' *z*-curve showing the 95% confidence interval for average statistical power was 5% to 13% and an observed discovery rate of 37% to 56%. Four recent pre-registered studies with relatively large sample sizes have all failed to observe a self-controlled practice or learning benefit (Bacelar, Parma, Cabral, et al., 2022; McKay & Ste-Marie, 2022; St.

Germain et al., 2022; Yantha et al., 2022). For example, Bacelar, Parma, Cabral, et al. (2022) had 100 participants practice a non-dominant arm bean bag tossing task with self-controlled augmented knowledge of results feedback and 100 participants practice the task without choice, and did not find a self-controlled learning advantage. Regarding enhanced expectancies, our RoMBA-PSMA estimate of $d = 0.26$ is consistent with Bacelar and colleagues' (2022) concern that their meta-analysis yielded an overestimated effect size ($g = 0.54$). Indeed, the present study uses new methods and different assumptions to correct for the bias that Bacelar and colleagues (2022) suspected in their previous study. Further, the Bayesian approach currently employed offers an updated belief in the likelihood of bias, heterogeneity, and the presence of an effect while considering plausible forms of selective reporting. Conversely, the frequentist approach employed previously offers a test of the null hypothesis that there is no effect of enhanced expectancies manipulations while assuming selective reporting is absent. The two studies offer valid answers to different questions. In addition, our z-curve result that enhanced expectancies are underpowered and subject to reporting bias is consistent with Bacelar and colleagues' findings of small sample sizes (median = 14/group) and small-study effects (i.e., significant funnel plot asymmetry). Estimates of reporting bias from the RoBMA-PSMA models were sensitive to the removal of two outliers, showing strong support for bias only when one or two outliers were included but not when both were removed. The evidence of reporting bias observed in the z-curve was not sensitive to removal of outliers and with the heterogeneity in the sample the z-curve may provide a better test of bias than RoBMA-PSMA. Considering these motivational factors combined, the z-curve result that studies supporting OPTIMAL theory are underpowered and subject to reporting bias is consistent with McKay and colleagues' (2022) meta-analysis that drew the same conclusion about self-controlled practice studies. Taken together, past and present results suggest that the our samples of studies demonstrating the benefit of enhancing learners' expectancies and giving them control over practice conditions presents a distorted reality due to reporting bias. These effects are not reliably different from zero, highlighting that the motivation pillar of OPTIMAL theory lacks evidential value.

Limitations

Our efforts to model selective reporting and adjust our parameter estimates accordingly are limited by the difficult nature of this task. We simply cannot know the extent of reporting bias in the extant literature, nor can we know the underlying mechanisms responsible for it. While our models correspond nicely to plausible selection processes, there are other possible mechanisms that could cause the data to fit our bias-correction models. For instance, it is possible researchers were able to intuit the size of the effects they would observe with their specific sample and paradigm, and they carefully adjusted their sample sizes based on these intuitions. If this was the case, we would expect large studies for small effects and small studies for large ones, consistent with the regression models included in the RoBMA-PSMA. Although we consider this unlikely—how could researchers have such a fine-grained understanding of the effects they are studying given the uncertainty in the literature?—the reader should be aware of the sensitivity of our models to assumptions about the underlying data generating process. The likely presence of heterogeneity in the enhanced expectancies literature suggests there is not one true effect, so individual studies may have been testing interventions with real benefits. However, this also means individual studies may have been testing interventions with real detriments as well, and we cannot discern which studies fall into which category.

Conclusion

Our analyses suggest a lack of evidence in support of enhanced expectancies and self-controlled practice as beneficial motor learning interventions. The lack of evidence supporting the motivational branch in OPTIMAL theory is not evidence that the predicted effects are absent. Indeed, even the null effects for a self-controlled practice benefit reported by pre-registered studies with large sample sizes (Bacelar, Parma, Cabral, et al., 2022; McKay & Ste-Marie, 2022; St. Germain et al., 2022; Yantha et al., 2022) are not conclusive that an effect is absent. Rather, these null effects leave open the possibility that the effect is very small and, thus, not detectable even with relatively large sample sizes (e.g., N = 200 as in Bacelar, Parma, Cabral, et al. (2022)). Motor learning researchers often study skills performed in sports, which are often *games of inches*, so very small effects may be of practical interest. Nonetheless, the field of motor learning is not past asking whether self-controlled practice and enhanced expectancies have any benefit, so concerns about estimating the magnitude of a potential benefit are premature (Simonsohn, 2015). Thus, we urge any motor learning scientist(s) interested in clarifying whether self-controlled practice or enhanced expectancies boost motor learning to address the problems of underpowered and overworked study designs (Lohse et al., 2016) and the reporting bias revealed in the present meta-analysis and those by McKay, Yantha, et al. (2022) and Bacelar, Parma, Murrah, et al. (2022).

There are multiple ways to increase statistical power, such as accounting for between-subjects variance by using a covariate, like pretest motor performance, in an ANCOVA design (Vickers & Altman, 2001), and/or increasing the number of pretest and posttest trials (Maxwell et al., 1991). Perhaps the most common and effective way to boost power is to increase sample size. Lakens (2022) describes several approaches for determining sample size, including conducting an *a priori* power analysis. A recent survey of three popular motor learning journals revealed a low prevalence (84/635 or 13% in McKay, Corson, et al., 2022) and low reproducibility (7/84 or 8% in McKay et al., in-press) of reported *a priori* power analyses. The usefulness of an *a priori* power analysis depends on reasonable assumptions about effect sizes. Assumptions about the effect size for self-controlled practice and enhanced expectancies studies should be based on the bias-corrected estimates found in the present study (self-controlled practice: d = .034; enhanced expectancies: d = 0.26) given the evidence of reporting bias. Researchers could also use the smallest effect size of interest (Lakens, 2022), but this effect may be even smaller than those from the meta-analyses, as noted earlier. Simonsohn (2015) described another approach to determine sample size for replication studies termed the ‘small telescopes’ approach. This approach recommends the replication sample size be 2.5 times that of the original sample. With questions surrounding the face value of original studies supporting OPTIMAL theory, we believe replication studies are crucial and the ‘small telescopes’ approach to determine sample size for these studies should be the minimally accepted approach. Irrespective of which of the above approaches researchers use to make reasonable assumptions about effect sizes, their *a priori* power analyses will likely lead to sample sizes that are larger than they are used to collecting. Thus, researchers may want to consider ways to improve the efficiency of their data collection, for example by using sequential analyses (Lakens et al., 2021; Lakens, 2014; Wald, 1945) or conducting multi-laboratory studies. Finally, clarity about the effect of self-controlled practice and/or enhanced expectancies on motor learning can only be achieved if a complete picture of the evidence is available. Researchers and gatekeepers to scientific publication (e.g., peer-reviewers, journal editors) should take measures to eliminate reporting bias, for example by publishing registered reports, undertaking/encouraging replication attempts (at a minimum using the 'small telescopes' approach), and publishing null effects.

Author Contributions (CRediT Taxonomy)

Conceptualization: [redacted]

Data curation: [redacted]

Formal analysis: [redacted]

Funding acquisition: [redacted]

Investigation: [redacted]

Methodology: [redacted]

Project administration: [redacted]

Software: [redacted]

Supervision: [redacted]

Validation: [redacted]

Visualization: [redacted]

Writing – original draft: [redacted]

Writing – review & editing: [redacted]

Data availability statement

The data and scripts can be accessed using either of the following links: https://osf.io/vfza7/ or [redacted].

Declaration of interest statement

All authors declare no conflicts of interest.

References

Abe, M., Schambra, H., Wassermann, E. M., Luckenbaugh, D., Schweighofer, N., & Cohen, L. G. (2011). Reward improves long-term retention of a motor memory through induction of offline memory gains. *Current Biology*, *21*(7), 557–562. https://doi.org/10.1016/j.cub.2011.02.030

Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://github.com/crsh/papaja

Ávila, L. T. G., Chiviacowsky, S., Wulf, G., & Lewthwaite, R. (2012). Positive social-comparative feedback enhances motor learning in children. *Psychology of Sport and Exercise*, *13*(6), 849–853. https://doi.org/10.1016/j.psychsport.2012.07.001

Bacelar, M. F. B., Parma, J. O., Cabral, D., Daou, M., Lohse, K. R., & Miller, M. W. (2022). Dissociating the contributions of motivational and information processing factors to the self-controlled feedback learning benefit. *Psychology of Sport and Exercise*, *59*, 102119. https://doi.org/10.1016/j.psychsport.2021.102119

Bacelar, M. F. B., Parma, J. O., Murrah, W. M., & Miller, M. W. (2022). Meta-analyzing enhanced expectancies on motor learning: Positive effects but methodological concerns. *International Review of Sport and Exercise Psychology*, *0*(0), 1–30. https://doi.org/10.1080/1750984X.2022.2042839

Barth, M. (2022). *tinylabels: Lightweight variable labels*. https://cran.r-project.org/package=tinylabels

Bartoš, F., & Maier, M. (2020). *RoBMA: An r package for robust bayesian meta-analyses*. https://CRAN.R-project.org/package=RoBMA

Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H., & Stanley, T. D. (2022). Robust bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*, 1–18. https://doi.org/10.1002/jrsm.1594

Bartoš, F., & Schimmack, U. (2020). *Zcurve: An r package for fitting z-curves*. https://CRAN.R-project.org/package=zcurve

Bartoš, F., & Schimmack, U. (2022). Z-curve 2.0: Estimating replication rates and discovery rates.

Meta-Psychology, *6*. https://doi.org/10.15626/MP.2021.2720

Cameron, A., & van den Brand, T. (2022). *Geomtextpath: Curved text in 'ggplot2'*. https://CRAN.R-project.org/package=geomtextpath

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. https://doi.org/10.1177/2515245919847196

Champely, S. (2020). *Pwr: Basic functions for power analysis*. https://CRAN.R-project.org/package=pwr

Chiviacowsky, S., & Wulf, G. (2007). Feedback after good trials enhances learning. *Research Quarterly for Exercise and Sport*, *78*(2), 40–47. https://doi.org/10.1080/02701367.2007.10599402

Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science*, *18*(4), 314–316. https://doi.org/10.1111/j.1467-9280.2007.01896.x

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 943–943. https://www.jstor.org/stable/24749235

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38. https://www.jstor.org/stable/2984875

Duval, S., & Tweedie, R. (1998). Practical estimates of the effect of publication bias in meta-analysis. *Australasian Epidemiologist*, *5*(4). https://doi.org/10.3316/informit.434005834350149

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLOS ONE*, *5*(4), e10271. https://doi.org/10.1371/journal.pone.0010271

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, *36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546–573. https://doi.org/10.1177/1745691616652873

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

Iannone, R., Cheng, J., & Schloerke, B. (2022). *Gt: Easily create presentation-ready display tables*. https://CRAN.R-project.org/package=gt

Kahle, D., & Stamey, J. (2017). *Invgamma: The inverse gamma distribution*. https://CRAN.R-project.org/package=invgamma

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*(7), 701–710. https://doi.org/10.1002/ejsp.2023

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Lakens, D., Pahlke, F., & Wassmer, G. (2021). *Group sequential designs: A tutorial*. PsyArXiv. https://doi.org/10.31234/osf.io/x4azm

Lee, C., Linkenauger, S. A., Bakdash, J. Z., Joy-Gaba, J. A., & Profitt, D. R. (2011). Putting like a pro: The role of positive contagion in golf performance and perception. *PLOS ONE*, *6*(10), e26016. https://doi.org/10.1371/journal.pone.0026016

Lee, G., & Scott, C. (2012). EM algorithms for multivariate Gaussian mixture models with truncated and censored data. *Computational Statistics & Data Analysis*, *56*(9), 2816–2829. https://doi.org/10.1016/j.csda.2012.03.003

Lee, T. D., & Carnahan, H. (2021). Motor learning: Reflections on the past 40 years of research. *Kinesiology Review*, *10*(3), 274–282. https://doi.org/10.1123/kr.2021-0018

Lewthwaite, R., Chiviacowsky, S., Drews, R., & Wulf, G. (2015). Choose to move: The motivational impact of autonomy support on motor learning. *Psychonomic Bulletin & Review*, *22*(5), 1383–1388. https://doi.org/10.3758/s13423-015-0814-7

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. https://doi.org/10.1123/jmld.2015-0010

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *00*, 1–20. https://doi.org/10.1002/ejsp.2752

Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2022). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*. https://doi.org/10.1037/met0000405

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

Maxwell, S. E., Cole, D. A., Arvey, R. D., & Salas, E. (1991). A comparison of methods for increasing power in randomized between-subjects designs. *Psychological Bulletin*, *110*(2), 328–337. https://doi.org/10.1037/0033-2909.110.2.328

McKay, B., Bacelar, M. F. B., & Carter, M. J. (in-press). On the reproducibility of power analyses in motor behavior research. *Journal of Motor Learning and Development*.

McKay, B., Corson, A., Vinh, M.-A., Jeyarajan, G., Tandon, C., Brooks, H., Hubley, J., & Carter, M. J. (2022). Low prevalence of a priori power analyses in motor behavior research. *Journal of Motor Learning and Development*, *1*(aop), 1–14. https://doi.org/10.1123/jmld.2022-0042

McKay, B., & Ste-Marie, D. M. (2022). Autonomy support via instructionally irrelevant choice not beneficial for motor performance or learning. *Research Quarterly for Exercise and Sport*, *93*(1), 64–76. https://doi.org/10.1080/02701367.2020.1795056

McKay, B., Yantha, Z. D., Hussien, J., Carter, M. J., & Ste-Marie, D. M. (2022). Meta-analytic findings in the self-controlled motor learning literature: Underpowered, biased, and lacking evidential value. *Meta-Psychology*. https://doi.org/10.15626/MP.2021.2803

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. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1, 1), 1–9. https://doi.org/10.1038/s41562-016-0021

Novomestky, F., & Nadarajah, S. (2016). *Truncdist: Truncated random variables*. https://CRAN.R-project.org/package=truncdist

Pedersen, T. L. (2022). *Patchwork: The composer of plots*.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, *95*(3), 355–386. https://doi.org/10.1037/0033-2909.95.3.355

Sanli, E., Patterson, J., Bray, S., & Lee, T. (2013). Understanding self-controlled motor learning protocols through the self-determination theory. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00611

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260. https://doi.org/10.1037/h0076770

Sievert, C. (2020). *Interactive web-based data visualization with r, plotly, and shiny*. Chapman; Hall/CRC. https://plotly-r.com

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results.

Psychological Science, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341

St. Germain, L., Williams, A., Balbaa, N., Poskus, A., Lohse, K. R., & Carter, M. J. (2022). Increased perceptions of autonomy through choice fail to enhance motor skill retention. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(4), 370–379. https://doi.org/10.1037/xhp0000992

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78.

Stefan, A., & Schönbrodt, F. (2022). *Big little lies: A compendium and simulation of p-hacking strategies*. PsyArXiv. https://doi.org/10.31234/osf.io/xy2dk

Ste-Marie, D., Vertes, K., Rymal, A., & Martini, R. (2011). Feedforward self-modeling enhances skill acquisition in children learning trampoline skills. *Frontiers in Psychology*, *2*. https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00155

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, *53*(11), 1119–1129. https://doi.org/10.1016/S0895-4356(00)00242-0

Sutton, A., Song, F., Gilbody, S., & Abrams, K. (2000). Modelling publication bias in meta-analysis: A review. *Statistical Methods in Medical Research*, *9*(5), 421–445. https://doi.org/10.1177/096228020000900503

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207–216. https://doi.org/10.1016/S0895-4356(99)00161-4

Ushey, K. (2022). *Renv: Project environments*. https://CRAN.R-project.org/package=renv

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ*, *323*(7321), 1123–1124. https://doi.org/10.1136/bmj.323.7321.1123

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, *16*(2), 117–186. https://www.jstor.org/stable/2235829

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, *23*(5), 1382–1414. https://doi.org/10.3758/s13423-015-0999-9

Wulf, G., & Lewthwaite, R. (2021). Translating thoughts into action: Optimizing motor performance and learning through brief motivational and attentional influences. *Current Directions in Psychological Science*, *30*(6), 535–541. https://doi.org/10.1177/09637214211046199

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

Yantha, Z. D., McKay, B., & Ste-Marie, D. M. (2022). The recommendation for learners to be provided with control over their feedback schedule is questioned in a self-controlled learning paradigm. *Journal of Sports Sciences*, *40*(7), 769–782. https://doi.org/10.1080/02640414.2021.2015945

Figure captions

Figure 1. Description of the original meta-analyses. Diagram outlining the shared and unique inclusion criteria, dependent variable selection, the number of studies screened, and the number of outcomes included in each meta-analysis.

Figure 2. Results from the robust Bayesian meta-analysis with publication selection model-averaging (RoBMA-PSMA) method. **(A)** Prior spike and distribution (purple) with 50% of prior probability density concentrated on the null hypothesis and 50% reflecting plausible true effects with $M = 0$ and $SD = 1$. The model-averaged posterior distribution of the mean effect (Mu) of **(B)** self-controlled practice on motor learning (blue). The increased height of the spike at mu = 0 reflects increased belief in the null hypothesis. The remaining distribution reflects updated belief in the size and direction of a possible true effect. The mean estimate for the posterior distribution is $M = .034$. The model-averaged posterior distribution of the mean effect of **(C)** enhanced expectancies on motor learning (red). The decreased height of the spike at mu = 0 reflects decreased belief in the null hypothesis. The remaining distribution reflected updated belief in the size and direction of a possible true effect. The mean estimate for the posterior distribution is $M = .26$. The model-averaged posterior distribution generated from an analysis of **(D)** simulated data (green). In the simulation, 49 studies were sampled from a population with a true effect of mu = .54 and no reporting bias. The mean estimate for the posterior is $M = .504$. Outliers were not included in the self-controlled ($n = 2$) and enhanced expectancies ($n = 2$) models.

Figure 3. Results of the *z*-curve analyses. Distribution of *z*-values for **(A)** our simulation with 47% results and no reporting bias (green), **(B, left)** self-controlled practice (blue), **(C, left)** enhanced expectancies (red), and **(D, left)** motivational factors combined (purple). Values in the *z*-score distributions for each analysis that are to the right of the significance line ($z = 1.96$; solid, black) are statistically significant with a two-tailed α of .05. Bootstrapped confidence distributions for the expected discovery rate (EDR; dark) and expected replication rate (ERR; light) for **(B, right)** self-controlled practice (blue), **(C, right)** enhanced expectancies (red), and **(D, right)** motivational

factors combined (purple). Note, reported confidence intervals include 5 extra points (EDR) and 3 extra points (ERR) added to the quantiles of the bootstrapped distributions in the right panel, consistent with Bartoš et al. (2022). The expected discovery rate is the estimated average power of all studies that have been conducted . The expected replication rate is the estimated power of all studies that observed a statistically significant result. The analysis estimated the average statistical power to be 6%, 8%, and 6% for self-controlled practice, enhanced expectancies, and motivational factors combined, respectively. The estimated conditional power of the statistically significant results was 11%, 33%, and 21% for self-controlled practice, enhanced expectancies, and motivational factors combined, respectively. Note that a minimum power of 80% (dashed line, black) is often recommended. Outliers were not included in the self-controlled (n = 2), enhanced expectancies (n = 2), and motivational factors combined (n = 4) analyses.