Educational Psychology Review (2025) 37:73
https://doi.org/10.1007/s10648-025-10043-1

REFLECTION ON THE FIELD

Advancing Contextual Interference: Addressing Methodological Debates, Reflecting on Meta-Analytic Practices and Generalizability, and Guiding Future Directions in Motor Learning

Achraf Ammar[1,2,3,5] · Khaled Trabelsi[3,4,5] · Atef Salem[1,3] · Haitham Jahrami[6,7] · Wolfgang I. Schöllhorn[1]

Received: 11 June 2025 / Accepted: 16 June 2025 / Published online: 24 July 2025

Abstract

Given that the contextual interference (CI) phenomenon is one of the most extensively studied and debated topics in motor learning—featured prominently in scientific literature, textbooks, and practitioner guides—it is unsurprising that recent meta-analyses on the topic have generated critical discussion and contrasting interpretations. This article responds comprehensively to the commentary by Czyż (2025) on our recent meta-analytic work evaluating CI effects across acquisition, retention, and transfer phases in sports related movements. We systematically address their methodological critiques, clarify inclusion decisions, and challenge inconsistencies in their own analytic approach and interpretation—particularly regarding baseline equivalence, study exclusion, and epistemologically inadmissible generalized claims. In addition to a detailed point-by-point rebuttal, we provide a comparative evaluation of recent CI-related meta-analyses and offer a broader reflection on the current state of meta-analytic practice in motor learning. We highlight concerns related to statistical ritualism, inconsistent quality control, and the growing redundancy of evidence syntheses. Finally, we present targeted recommendations for future research, including the need for high-quality randomized controlled trials, field-specific quality appraisal tools, and comprehensive umbrella reviews to better integrate fragmented evidence. Collectively, these reflections aim to enhance methodological standards and ensure more context-sensitive and scientifically sound conclusions in CI and broader motor learning research.

Achraf Ammar, Khaled Trabelsi, Haitham A Jahrami and Wolfgang I. Schöllhorn contributed equally as first authors.

Extended author information available on the last page of the article

Keywords Interference effect · Random practice · Serial practice · Blocked practice · Applied setting · Acquisition · Retention · Transfer · Sport · Motor skills · Meta-analysis · Systematic review

Introduction

Meta-analysis is a cornerstone statistical method used to synthesize research findings from multiple individual studies within a given field (Hansen et al., 2022). Its primary objectives are to enhance statistical power, provide more accurate estimates of effect sizes, and help resolve discrepancies when individual studies report conflicting results (Gurevitch et al., 2018). Meta-analysis is also a valuable tool for testing competing theoretical assumptions and/or identifying key moderators (Aguinis et al., 2011; Bergh et al., 2016). Beyond summarizing existing evidence and integrating extant research findings, meta-analyses also guide future research by highlighting gaps in the literature and suggesting promising avenues for further investigation (Chan & Arvey, 2012).

While meta-analyses have the potential to improve precision, help deciding about questions not posed by individual studies, and help resolve scientific debates, its conclusions may also be misleading if methodological quality, study design differences, within-study biases, or publication and reporting biases are not adequately addressed (Deeks et al., 2023). To mitigate these risks, meta-analytic methodologies continue to evolve—particularly in fields like sport science, where heterogeneity is high and individual responses to interventions can vary substantially (Afonso et al., 2024; Steele et al., 2023). Since its origins in the 1970 s for evaluating the efficacy of medical and psychological interventions (Glass, 2015; Gurevitch et al., 2018), meta-analysis has become a well-established approach for synthesizing evidence across numerous disciplines, including motor learning (McKay et al., 2022a, 2022b), and specifically the contextual interference (CI) effect.

Several meta-analyses have already examined the CI hypothesis, which posits that high CI (e.g., random practice) leads to inferior acquisition (1st effect), but superior learning capabilities (2nd effect), operationalized in form of retention and/or transfer tests, compared to low CI (i.e., blocked practice). Brady (2004) was a pioneer in conducting a basic meta-analysis on CI, comparing cumulative effect sizes across applied and laboratory contexts as well as across age groups. However, his work addressed only the second CI effect. Sattelmayer et al. (2016) later examined both CI effects, but their analysis was limited to studies in physiotherapy and medical education.

Given the growing body of sport-specific motor learning studies and advances in meta-analytic methods since Brady's work, Ammar et al., (2023, 2024) conducted two comprehensive meta-analyses to address the mixed findings surrounding both CI effects. These analyses focused on gross motor skills and evaluated performance across key time points—post-test, retention, and transfer—either by examining outcome measures at each stage (Ammar et al., 2023) or by analyzing computed changes across phases to capture both acquisition and relatively permanent gains (Ammar & Schöllhorn, 2024). Collectively, these meta-analyses found insufficient evidence to support generalizable recommendations in favor of high CI, particularly in applied settings.

In 2024, Czyż et al. published two additional meta-analyses on the effects of high contextual interference in motor learning. The first focused on retention performance (Czyż et al., 2024a), while the second addressed transfer (Czyż et al., 2024b). Although both meta-analyses included studies on fine and gross motor tasks, they were limited to analyzing the second CI effect—retention or transfer—while neglecting the initial interference effect on acquisition performance (Czyż et al., 2024a, b). We read both studies with great interest given their thematic overlap with our work and have previously highlighted some concerns in our 2024 commentary (Ammar & Schöllhorn, 2024), particularly regarding the first study, which made epistemologically inadmissible generalized claims in favor of the CI effect—as reflected in its title: "High contextual interference improves retention in motor learning: systematic review and meta-analysis." Although these meta-analyses were essentially an update of Brady's (2004) work, they continued the same narrow focus on the second CI effect (retention) while overlooking the first acquisition effects. Despite including a majority of low-quality studies, potential risk of publication bias, and acknowledging that CI benefits are nearly negligible in applied settings (as noted by the authors themselves in the abstract: "On the contrary, in the applied setting, the beneficial effect of random practice on retention was almost negligible"), the paper concluded with a strong and inadmissible generalized claim: "The CI effect is a robust phenomenon in motor learning." Such a conclusion, especially when not qualified (e.g., "in applied settings"), risks being misleading. Given the ongoing challenges highlighted by the replication crisis in behavioral sciences, such generalizations warrant greater caution and nuance. Accordingly, in our previous commentary (Ammar & Schöllhorn, 2024), we emphasized that these meta-analyses should be interpreted with caution, and that broad, definitive conclusions should be avoided given the limitations acknowledged within the study itself.

Given that CI phenomenon is one of the most widely explored topics in motor learning literature, including textbooks and practitioner guidelines—and considering the inconsistent findings across meta-analyses, particularly those revealed by Ammar et al. and Czyż et al. team—it was not unexpected that our recent meta-analysis would spark discussion and debate. In this context, we appreciate the commentary by Czyż (2025) on our critical review of CI effects on acquisition and relatively permanent gains (Ammar et al., 2024), which was also published in the prestigious Educational Psychology Review, known for its rigor and adherence to best scientific practices.

In response to the concerns raised about our meta-analysis—and given that Czyż et al. extended their critique beyond Ammar et al. (2024) to also question aspects of Ammar et al. (2023), though not all points were accurate (e.g., claims regarding the absence of sensitivity analyses)—the present paper systematically addresses these issues through a detailed point-by-point rebuttal. Where appropriate, we also provide counterexamples drawn from Czyż et al.'s own recent meta-analyses (2024a, b) (see the "Point-by-Point Rebuttal to Czyż Commentary on Ammar et al. (2024)" section).

In the "Brief Comparative Evaluation of Recent CI Related Meta-analysis and Reflection on Meta-analytics Practice in Motor Learning" section, we provide a brief evaluation of these recent meta-analyses, followed by a broader reflection on current meta-analytic practices in the motor learning field. We highlight key concerns around generalizability and the increasing production of systematic reviews and meta-analyses in the absence of sufficient high-quality randomized controlled trials—an issue raised by Ioannidis (2016), who warned of the risks of unnecessary, redundant, or potentially biased syntheses. We also reflect on the evolving methodologies in meta-analysis and their implications for evidence synthesis in motor learning research.

Finally, in the "Future Directions in Contextual Interference and the Broad Motor Learning Research" section, given the generally low methodological quality of original studies in the CI literature (as noted in Ammar et al., 2023, 2024; Czyż et al., 2024a, b), the proliferation of recent meta-analyses, and the substantial heterogeneity in motor learning research, we offer reflections on future directions in the field. We emphasize the urgent need for high-quality, well-powered randomized controlled trials that emphasize the simultaneous comparison of multiple motor learning approaches and advocate for a shift toward comprehensive umbrella reviews to evaluate and synthesize existing meta-analyses—rather than producing further potentially unnecessary or duplicative meta-analyses.

Point-by-Point Rebuttal to Czyż, (2025) Commentary on Ammar et al. (2024)

Database Selection and Potential Omitted Studies

We thank the authors for their thoughtful examination of our search strategy and for raising points that can help guide future systematic reviews. Below, we provide clarifications and counterpoints to demonstrate that our approach was methodologically sound, transparent, and aligned with current best practices. Our systematic review employed a multi-database search using five reputable and multidisciplinary databases: PubMed, Scopus, Web of Science, SciELO, and Taylor & Francis. This selection ensured broad coverage across biomedical, psychological, sports science, and educational research domains. As emphasized in the PRISMA 2020 guidelines (Page et al., 2021; Rethlefsen et al., 2021), a robust search strategy should aim for a balance between comprehensiveness and feasibility, tailored to the research question and the disciplinary landscape. In this regard, our strategy aligns and even exceeds meta-research recommendations, such as those by Ewald et al. (2022), who concluded that using two or more databases significantly improves study identification and minimizes publication bias.

Moreover, we adhered to PRISMA's recommendation to balance sensitivity and specificity in literature searches. Rather than attempting exhaustive inclusion from every possible source, we focused on broad, multidisciplinary databases that consistently index peer-reviewed, high-quality research relevant to motor learning, sport science, and psychology. The combined use of Scopus and Web of Science, with their substantial overlap and citation depth, further ensured a high capture rate of pertinent literature while minimizing redundancy and improving screening efficiency. Importantly, PRISMA also underscores the need to consider disciplinary relevance, research design, and question specificity when selecting databases. Our review centered on gross motor skill acquisition and retention under CI conditions, a topic commonly addressed through experimental designs in sports and movement science—areas well represented in the selected databases. The inclusion of PubMed allowed us to access medically indexed studies, while Taylor & Francis and SciELO enhanced regional and open-access inclusivity, ensuring representation of studies beyond typical Western publication circuits.

In general, our database selection was intentionally broad, strategically reasoned, and aligned with PRISMA guidelines and best practices in systematic reviewing. The methodological rigor of our approach supports both the integrity of our findings and the reproducibility of our search process.

Specifically, while domain-specific databases such as SPORTDiscus and PsycINFO (highlighted in the current commentary) may enhance coverage in some cases, it is important to also note the following:

Significant Overlap Exists Across Databases

Studies indexed in SPORTDiscus and PsycINFO are frequently retrievable via PubMed, Web of Science, or Scopus. Our search successfully captured numerous key motor learning and sport psychology studies (e.g., French et al., 1990; Hall et al., 1994; Pollatou et al., 1997), confirming effective coverage of the field.

Omitted Studies Likely Had Limited Impact on Overall Findings

The commentary highlights three studies—Jiménez-Díaz et al. (2018), Moretto et al. (2018), and Tsutsui et al. (2013)—as missing from our synthesis. While these studies were not retrieved during our original search, we have reviewed them post hoc and provide the following clarifications:

(i) Jiménez-Díaz et al. (2018): This study reported no significant differences between blocked and random practice schedules, concluding that contextual interference had no effect on fundamental motor skills in adults. If retrieved and included, it would have aligned with—and potentially reinforced—our overall conclusion of no consistent CI advantage in this domain.

(ii) Moretto et al. (2018): Based on available information, this study does not meet our methodological inclusion criteria, particularly regarding the absence of baseline equivalence between groups. As detailed in our protocol, baseline balance was a core inclusion criterion. Therefore, even if retrieved, it would have been excluded from our analysis.

(iii) Tsutsui et al. (2013): This study involved extremely small sample sizes ($n=5$ per group) and reported mixed findings across high and low skill levels. Given its limited statistical power and inconsistent results, its potential impact on a meta-analysis including over 1100 participants would likely have been negligible.

We acknowledge that no systematic review is immune to missing a few eligible studies. The key question is whether such omissions substantially affect the findings. In this case, based on the studies' content, sample sizes, and alignment with our conclusions or exclusion criteria, we are confident that their absence did not meaningfully alter the outcomes or interpretations of our meta-analysis.

Commentary Authors Omitted Major Studies and Databases

Notably, the authors of the current commentary (Czyż, 2025) conducted two follow-up meta-analyses on contextual interference—one on retention (Czyż et al., 2024a) and another on transfer (Czyż et al., 2024b)—shortly after our work was published, yet their database search omitted PubMed, a primary source for medical and sports science research. Although EBSCO includes MEDLINE content, PubMed provides broader access, encompassing PubMed Central, in-process records, and ahead-of-print citations, thereby offering more comprehensive coverage.

Moreover, our review included several "applied setting" studies that—despite aligning with the scope and PICOS criteria of Czyż et al.'s systematic reviews (2024a, b)—were either not retrieved or were excluded without convincing justification. It is important to note that we limited our screening to applied-setting studies; thus, if this assessment were extended to laboratory-based studies, the number of potentially relevant but omitted articles in their reviews would likely be considerably higher. Examples of such overlooked applied-setting studies include: (i) Boyce and Del Rey (1990): This study was not referenced in their systematic reviews or appendices. In their recent rebuttal to our commentary (Czyż & Wójcik, 2024), they argued that the full text could not be retrieved. However, their original review stated that "titles and abstracts of 1,255 articles were screened according to PICO criteria, of which 961 records were excluded due to topic reasons, study design issues, and population," with no mention of exclusion due to unavailable full texts. The rationale for omitting this study thus remains unclear and inconsistent with their stated criteria; (ii) Memmert (2006): In their rebuttal (Czyż & Wójcik, 2024), the authors claimed this study was excluded during the abstract screening phase because it involved constant rather than blocked practice. However, their own PICOS criteria listed "blocked schedule/repetitive practice" as the control condition. A study using repeated execution of a single task should therefore qualify as "repetitive practice," which falls under low contextual interference. Further justification is provided in our response to Point 3; (iii) Rendell et al. (2010): The authors stated that this study was excluded due to missing mean and SD values. However, the original paper presents results graphically. These values can be accurately extracted using tools such as WebPlotDigitizer (e.g., Pacifica, CA, USA, Version 4.5), as was done by Ammar et al. (2023), particularly in cases where raw data were not available upon request; (iv) North et al. (2019): In Appendix 2, the study was excluded on the basis that only difference scores (pre-test to retention) were reported and raw retention test results were unavailable. Upon consulting the full text, however, retention data are clearly presented in graphical form, making them retrievable using the same approach noted above; (v) Jones and French (2007): This study was excluded on the basis that the exact number of participants per group was not provided—only the total number "51," despite a clear indication that participants were randomized across three groups. However, this limitation is not critical, particularly for inclusion in the systematic review. Nonetheless, the study was excluded from both the systematic review and the meta-analysis without further justification; and (vi) Sadri et al. (2013): This study was excluded for "unsatisfactory quality." However, since the review aimed to assess and report study quality, such a criterion should be used for grading—not excluding—studies. Moreover, no detailed justification was given to warrant exclusion from both the systematic review and the meta-analysis.

These examples highlight methodological issues that we believe deserve equal scrutiny alongside those raised about our work. Notably, many of the omitted studies reported no CI benefit at retention (e.g., Boyce and Del Rey (1990); Memmert (2006); Jones and French (2007); Sadri et al. (2013)) and/or at transfer (e.g., Memmert (2006); Rendell et al., (2010)) in the majority of tested skills or conditions (Ammar et al., 2023). This calls into question the consistency and robustness of the beneficial CI effects reported in their meta-analyses.

Additionally, several studies included in their reviews—such as Smith and Rudisill (1993), Smith and Davies (1995), Li and Lima (2002), Brady (1998), and Fialho et al. (2006)—did not ensure group equivalence at baseline, introducing potential risk of bias into their comparisons of retention or transfer outcomes. Ammar et al., (2023, 2024) explicitly argued that baseline equivalence between groups is a critical methodological prerequisite for attributing learning effects to the intervention itself, rather than to pre-existing differences in performance. This consideration is especially important in motor learning research, where higher levels of participant expertise often require disproportionately greater effort to achieve comparable learning gains.

Our meta-analysis strictly followed this principle, excluding 25 studies lacking verified baseline comparability. Including such studies, as in Czyż et al., (2024a, b), requires an analytic strategy that does not rely solely on post-test comparisons but instead accounts for pre-post deltas or retention gains relative to acquisition—an approach we implemented in our analyses, but which is notably absent in both meta-analyses conducted by Czyż et al.,(2024a, b).

While Czyż & Wojcik (2024) acknowledge, in response to Ammar & Schöllhorn (2024), that baseline imbalances can influence pooled effect estimates, their analysis does not implement the specific strategies needed to control for this bias. In an attempt to argue their approach, Czyż & Wojcik (2024) cited Cochrane and Deeks et al. (2023) to justify including studies with potential baseline imbalance, provided this is addressed through sensitivity analysis and risk of bias assessments. Although this citation is accurate in principle, their response appears more theoretical than operational, as no analysis directly accounts for baseline group equivalence was performed in any of their meta-analyses, leaving several limitations in how this issue was actually handled in their meta-analysis such as:

1. Lack of sensitivity analysis focusing on baseline imbalance

   Although the authors reference the need for sensitivity analysis (Czyż & Wójcik, 2024), they did not report or perform themselves any subgrouping focused sensitivity analysis. The authors conducted a sensitivity analysis using Cook's D to identify influential outliers, but this approach helps address statistical outliers, but it does not identify design-related biases such as unequal pre-test performance. The decision to retain or exclude studies with baseline discrepancies is not evaluated or tested in both meta-analyses. Thus, their sensitivity analysis does not address the “baseline equivalence” concern raised by Ammar & Schöllhorn (2024).

2. No subgroup stratification based on baseline comparability

While heterogeneity was addressed using statistical models and sub-analyses, no one appear to account specifically for the equivalence of baseline data—a critical omission given its direct relevance to outcome interpretation. Furthermore, contrary to their own citation of Deeks et al. (2023) and Cochrane guidelines, the authors did not exclude or clearly downgrade studies that failed to report or control baseline performance (e.g., re-analysis excluding studies with missing or unreported pre-test data). This omission weakens the robustness of pooled post-test comparisons, which may reflect baseline disparities rather than true effects of contextual interference. Post-test comparisons without baseline control are vulnerable to bias, and failure to stratify by this criterion leaves conclusions uncertain.

3. Assumption that randomization ensures equivalence

Czyż & Wojcik (2024) appear to assume that because “randomization” was considered in their risk of bias tool, the risk of baseline differences is sufficiently controlled. However, randomization does not ensure baseline equivalence, particularly in small-sample studies common in motor learning research. Without reporting or statistically verifying pre-test equivalence, this remains an untested assumption.

4. Non-specific risk of bias application

The use of the Effective Public Health Practice Project Quality Assessment Tool (EPHPP) includes a general “confounders” category, but their reviews do not report how studies were scored specifically on baseline comparability and there is no discussion of how baseline comparability would influence results. No link is made between risk of bias ratings and pooled estimates—leaving a key confounding variable a potential uncontrolled bias.

5. Elevated risk in CI research

Given the small sample sizes and variability in CI literature, baseline differences are especially impactful. Studies with unbalanced pre-tests can easily skew acquisition, retention and/or transfer comparisons, overstating the benefits of random practice. Failure to account for this bias poses a serious threat to internal validity.

Taken together, while Czyż & Wojcik (2024) recognize the risk of baseline imbalance, their current analytic approach does not substantively address this threat. The absence of subgrouping, stratification, or sensitivity analysis related to baseline comparability limits confidence in the pooled effect estimates. Consequently, the core concern raised by Ammar et al.—that unequal pre-test performance may inflate

CI effects—remains unresolved and their conclusions—particularly the reported CI benefits—remain open to question. Specifically, their pooled effects may overestimate the benefits of CI, and conclusions drawn from such analyses should be interpreted with caution. In contrast, our reviews applied rigorous exclusion criteria to ensure that comparisons were methodologically sound and not confounded by pre-existing group differences.

For future research in this area, we encourage future original studies to account for baseline equivalence while we recommend reviews to adhere to more rigorous inclusion criteria or conduct stratified/sensitivity analyses specifically designed to account for baseline imbalances in order to provide more valid, less biased and interpretable effect estimate.

Comparative Transparency and Supplementary Material

Ammar et al. (2024) study provides full PRISMA flowcharts, a detailed table of included studies, and an appendix listing all excluded full texts alongside their justifications (Appendix S2). Such transparency ensures reproducibility and invites constructive critique. We note that this level of disclosure is not fully matched in the corresponding Czyż et al., (2024a, b) meta-analyses, particularly with regard to the majority of omitted relevant studies mentioned in the "Commentary Authors Omitted Major Studies and Databases" section.

In summary, regarding the discussion on database selection and potentially omitted studies, we fully agree that future updates and replications should consider incorporating additional sources such as SPORTDiscus and PsycINFO to further enhance completeness. Nonetheless, we respectfully affirm that our database selection strategy was methodologically sound, transparently documented, and aligned with current best practices in systematic reviews. As detailed above, we also affirm that the few studies potentially omitted did not substantially affect the results or interpretations of our meta-analysis.

Pre-registration

We thank Czyż (2025) for raising the issue of pre-registration. We fully acknowledge that registering a systematic review protocol prior to data collection is, in principle, a valuable scientific practice. It is increasingly recognized as a standard that promotes transparency, reduces bias, and enhances reproducibility. However, empirical evidence suggests that the practical impact of pre-registration may not fully meet these expectations. For example, a recent comparison of 193 pre-registered and 193 non-preregistered studies found no significant differences in the proportion of positive results, effect sizes, or statistical error rates (von den Akker et al., 2024). This may help explain why the PRISMA 2020 guidelines encourage registration but do not consider it a mandatory criterion for scientific validity. For a more detailed and critical discussion of pre-registration as a potentially ritualized practice, see also Pham & Oh (2021).

Indeed, as the authors themselves note, “pre-registration is not universally required in all fields.” It is particularly important to distinguish between recommendation and requirement. In our case, although we did not pre-register our protocol, our methodology was clearly documented, systematically applied, and fully reported. The search strategy, inclusion/exclusion criteria, data extraction, and analysis plan were pre-defined and followed rigorously, as detailed in our methods section and supplementary materials.

However, while Czyż (2025) offers a general case for the value of preregistration—most of which we agree with—their critique shifts into speculative territory when applied specifically to Ammar et al. (2024). For instance, they write:

> “Without a pre-registered protocol, researchers may be tempted—consciously or unconsciously—to adjust inclusion criteria or selectively report outcomes after seeing the data...”

They also suggest that “material or intellectual conflict of interest” may drive decisions such as which tests to include, what databases to use, or which studies to cite—thereby implying selective reporting or HARKing in our meta-analysis. We do not find sufficient evidence to support this implication in the context of our methodology.There is no evidence that Ammar et al. (2024) deviated from pre-specified inclusion criteria or adapted the analysis in response to findings. On the contrary, the study followed PRISMA recommendations, used a transparent, multi-database strategy (PubMed, Scopus, Web of Science, SciELO, and Taylor & Francis), and included extensive quality checks, subgroup analyses, and exclusion transparency (Appendix S2).

Moreover, it is important to note that while the commentary is clearly directed at Ammar et al. (2024), the specific example used to illustrate the risks of lacking pre-registration refers instead to a different study—Ammar et al. (2023). This undermines the relevance of the critique, as no concrete claim is made about how preregistration would have altered the design, conduct, or findings of the actual target paper (Ammar et al., 2024). Even for the 2023 study, the authors raise concerns about potential post hoc decisions (e.g., categorization of motor skills or retention intervals), but provide no evidence that such decisions were applied inconsistently or in a biased manner. If pre-registration is to be invoked as a key limitation, then it is incumbent upon the critics to specify which methodological components appear influenced by post hoc reasoning—something the commentary fails to do. As such, the argument would benefit from further substantiation in relation to our study.

Notably, while questioning our transparency, Czyż et al.’s own approach raises methodological issues on this very matter. Their meta-analyses (Czyż et al., 2024a, b) were based on six databases: Scopus, EBSCO, Web of Science, PsycINFO, ScienceDirect, and Google Scholar. However, when cross-checking these databases against key applied-setting studies omitted from their review, we found that several relevant works (e.g., Memmert (2006), Rendell et al. (2010), North et al. (2019), Jones and French (2007), Sadri et al. (2013)) are indeed retrievable via these databases. This indicates that the omission of these studies was not due to search coverage but rather their exclusion—deliberate or not—despite these studies being aligned with the pre-registered (Prospero ID: CRD42021228267) inclusion/exclusion criteria (as we explained in point 1.2) and highly relevant to the research question.

Importantly, a common pattern among these omitted studies is that many reported null or negative findings regarding CI effect. These patterns raise questions about potential unintentional selection effects that merit further clarification.

While Czyż et al., (2024a, b) did pre-register their protocol, the presence of a pre-registration does not automatically guarantee transparency in execution. What remains unclear—and unreported—is whether the studies we identified were retrieved during their screening phase and, if so, why they were excluded. No justification is provided in their published work or protocol regarding the omission of several relevant applied-setting studies that are retrievable via the databases they used. This lack of clarification undermines the benefits of pre-registration, which relies on consistent application and transparent reporting of inclusion/exclusion decisions.

In this context, the absence of justification for excluding multiple accessible and relevant studies—most of which challenge the narrative of high CI benefit—represents a tangible risk to transparency and bias mitigation. This is exactly the type of issue that pre-registration is designed to prevent. The claim that Ammar et al. (2024) may have selectively reported outcomes (despite no evidence of such practice) would more credibly apply to Czyż et al., (2024a, b), where omissions appear more systematic and not fully transparent. Without documentation of their exclusion decisions or subgroup rationale (especially for applied settings), one could argue that their process may have greater implications for internal validity than the hypothetical risk they attribute to ours. Moreover, when assessing the quality of the reviews by Czyż et al., (2024a, b) using the AMSTAR-2 (A MeaSurement Tool to Assess Systematic Reviews) (Shea et al., 2017), the lack of justification for excluded studies was identified as a critical weakness. This omission compromises the transparency of the review process and contributes to a rating of very low overall confidence in the findings.

In summary, we acknowledge the added value of preregistration, and future work should strive to include it. However, the absence of preregistration does not automatically imply flawed practice—especially in a study that provides complete methodological transparency and aligns with PRISMA 2020. We consider the criticisms of Ammar et al. (2024) regarding potential post hoc adjustments yet to be empirically demonstrated. In contrast, although Czyż et al., (2024a, b) did pre-register their protocol, the omission of several eligible applied-setting studies—without full reported justification—raises concerns about the consistent application of their criteria. These omissions, despite accessible sources and alignment with their stated inclusion rules, may have greater implications for internal validity than the hypothetical risks attributed to our review.

We welcome the growing emphasis on pre-registration but emphasize that its purpose is to enhance transparency, not to serve as a post hoc disqualifier when the absence of evidence-based misconduct exists. We remain committed to advancing transparency and reproducibility and encouraging future systematic reviews in the motor learning field to both pre-register and openly justify study selection—especially when dealing with hypotheses of high theoretical and applied significance.

Pre-registration is recommended—but not required—under PRISMA guidelines. Despite not pre-registering, we adhered to rigorous methodological standards, applied clearly defined inclusion criteria, and transparently justified all study selection decisions. The risk of bias due to HARKing (hypothesizing after results are known) was minimized through independent dual screening, structured data extraction, and specified sensitivity analyses. Contrary to the speculation raised by Czyż (2025), no post hoc modifications were made to our inclusion criteria after examining study outcomes. In contrast, we encourage the authors to clarify the rationale behind the omission of several relevant applied-setting studies from their own reviews—particularly given that many of these are accessible through their selected databases and appear to meet standard eligibility criteria.

Study Inclusion Criteria and Consistency

We thank the authors for raising concerns related to study classification in our meta-analysis. However, we respectfully disagree with their assertion that the inclusion of Memmert (2006) and Menayo et al. (2010) reflects inconsistencies with our stated inclusion criteria or could affect the validity of our analysis. Accordingly, we offer the following justifications for the inclusion of the studies in question, based on our pre-specified PICOS framework, the broader scope of CI conceptualization, and the empirical and theoretical basis for our classification decisions.

Regarding the inclusion of Memmert (2006), its inclusion was justified within the broader framework of CI, which encompasses comparisons between practice structures that differ in task variability and organization. While the commentary authors argue that this study did not involve a “blocked” condition in the strictest sense, it is important to recognize that the definitions of “blocked” and “random” practice vary across the literature. In our review, we adopted a theory-driven, inclusive classification of CI—supported by early foundational work such as Lee and Magill (1983)—that includes comparisons between constant (i.e., low CI) and variable/random (i.e., high CI) practice formats. The core principle of CI lies in the degree of cognitive challenge induced by differing practice schedules. While traditional CI paradigms typically compare blocked and random practice sequences involving multiple tasks (e.g., A-B-C), the broader conceptualization of CI—as articulated by Lee and Magill (1983) and supported by the elaboration and reconstruction hypotheses—recognizes task variability and contextual change as central learning mechanisms. In this view, constant practice of a single variation represents minimal interference (low CI), while variable practice involving non-repeating variations introduces high contextual demand (high CI). This interpretation is not unique to our review. Within CI theory (Battig, 1966; Shea & Morgan, 1979), repetitive practice of a single task—like blocked practice of multiple tasks—is recognized as inducing low contextual interference. Numerous meta-analyses and experimental studies (e.g., Brady, 2004; Sattelmayer et al., 2016) have treated repetitive, single-task conditions as valid low

CI control groups, particularly in single-skill learning contexts. Notably, even the authors of the commentary (Czyż & Wójcik, 2024) explicitly included "repetitive practice" in the control condition definition of their own PICOS criteria. Therefore, the inclusion of Memmert (2006) in our analysis aligns with established theoretical frameworks and is consistent with precedent in the peer-reviewed literature.

In the specific case of Memmert (2006)'s study, the variable group practiced different shot types from various locations in a varied sequence, whereas the constant group repeated the same shot from a fixed location. This design clearly compares a high contextual demand condition with a lower one—meeting the theoretical criteria for CI. Moreover, Memmert's study measured both acquisition and delayed retention, aligning with our primary outcome domains. Our goal was to provide an ecologically valid and comprehensive synthesis of the effects of high versus low contextual interference across a range of gross-motor specific interventions. Our PICOS framework was carefully designed to accommodate a broad spectrum of CI-relevant paradigms, provided that key methodological standards (e.g., baseline equivalence, retention or transfer measurement) were met. Memmert (2006) fulfilled these conditions and was thus appropriately included.

Regarding the inclusion of Menayo et al. (2010), the claim that Menayo et al. (2010) implemented a "simultaneous treatment design" that invalidates the distinction between blocked and random conditions may be misleading. Although the study reports that "the first 48 repetitions of each shot were executed in blocked practice, and the 144 remaining shots were executed in a random sequence," it also clearly states earlier in the methods section that "players were randomly assigned to the two practice conditions (blocked and random)" and that "during the sessions, each tennis player executed a total of 192 repetitions of shots, distributed in blocked or random practice." These statements collectively indicate that participants were allocated into distinct blocked and random conditions. Coupling these statements, with the repeated emphasis on condition-based randomization suggests that the data collection and analysis were structured around condition comparisons. This interpretation aligns with our inclusion criteria, which requires a comparison between higher and lower contextual interference conditions.

In addition, our broader aim was to synthesize findings from ecologically valid, applied settings to better understand how variations in practice structure influence motor learning. Within this context, Menayo et al.'s acquisition and retention comparisons between practice conditions provided relevant data and were thus considered appropriate for inclusion.

It is also important to note that our coding protocol and moderator analyses were designed to address potential variability in study design and execution. We systematically extracted and analyzed key methodological features—including study design type, skill classification, number of acquisition sessions, number of acquisition trials, and timing between acquisition and outcome assessments. These variables were incorporated into meta-regression and subgroup analyses to evaluate whether methodological differences are potentially moderate CI effects. The tested moderators captured critical dimensions of protocol heterogeneity relevant to CI interpretation.

Concerning the allegation of threatened validity, we respectfully disagreethe accusation that including Memmert (2006) and Menayo et al. (2010) raises concerns about the validity of our meta-analysis. First, as explained above, their inclusion was consistent with both our stated eligibility criteria and with established theoretical frameworks underpinning CI. Second—and more importantly—the direction of findings in these studies was inconsistent and non-systematic. For example, in the retention phase of Menayo et al., the performance advantage between high and low CI conditions was relatively balanced across comparisons. This lack of a consistent pattern suggests that their inclusion did not systematically favor one condition over the other and, therefore, does not compromise the integrity or interpretive value of our findings. Excluding these studies solely based on overly rigid interpretations of CI definitions would have unnecessarily narrowed the scope of our analysis and limited its ecological validity. On the contrary, their inclusion helped represent the broader range of designs employed in applied sport contexts and contributed to a more comprehensive and representative synthesis of the literature.

In summary, our classification strategy was transparently documented, theoretically grounded, and appropriate given the heterogeneity inherent in the CI literature. The inclusion of Memmert (2006) and Menayo et al. (2010) adhered to our inclusion criteria and reflected our commitment to synthesizing evidence across diverse, real-world training environments. Finally, we emphasize that even if these two studies had been excluded, their mixed and non-directional findings would not have meaningfully altered the primary results or overall conclusions of our meta-analysis. Thus, we do not consider their inclusion to compromise the validity of our analysis, and we stand by the robustness and interpretive integrity of our findings.

Paradoxically, even in their critique of classification consistency, Czyż et al., (2024a, b)'s own methodological choices reveal a degree of definitional flexibility. First, in their meta-analyses, they explicitly defined the control condition as "blocked schedule/repetitive practice." This broad definition clearly indicates that the commentary authors themselves recognized two forms of low contextual interference: (i) blocked practice—involving the repeated performance of task A before switching to B—and (ii) repetitive practice—involving the exclusive practice of a single task throughout the session. This definition contradicts the stricter interpretation they now apply in their commentary on our meta-analysis, thereby calling into question their rationale for excluding studies such as Memmert (2006) from their own analyses. In contrast, our inclusion of such studies is consistent with both the definitions they previously endorsed and broader precedent in CI literature. Their revised stance reflects either an inconsistency or a post hoc change in inclusion criteria—one that, at a minimum, requires transparent justification.

Second, their inclusion of two studies by Beik and Fazeli (2021) and Beik et al. (2021)—each featuring six groups that varied by both practice schedule (blocked, random, algorithm-based) and task similarity (similar vs. dissimilar)—raises further questions. While they appropriately excluded the algorithm-based groups, they included both similar- and dissimilar-task comparisons under blocked and random formats without clarifying the theoretical implications of this distinction. The decision to categorize tasks based solely on phenomenological similarity is problematic, as it represents a weaker operationalization of the CI hypothesis and reflects a common conceptual limitation in CI research (Buszard et al., 2017; Schöllhorn, 2016). The CI framework traditionally differentiates two practice conditions even within the random condition based on the manipulation of generalized motor programs (GMPs), primarily defined by invariant features such as relative timing and relative forces (Wulf and Lee, 1993). Notably, in studies involving fine motor tasks, greater CI effects are expected when switching between different GMPs (Wulf and Lee, 1993), whereas in gross motor tasks, CI benefits are more likely to emerge when variability occurs within the same GMP (Apidogo et al., 2021, 2022). While the former is typically attributed to increased cognitive load from task switching, the underlying mechanism for the latter remains under investigation. Including task conditions with unclear or mixed GMP classifications—while simultaneously criticizing other studies (e.g., Memmert, 2006; Menayo et al., 2010) for allegedly deviating from CI norms—reflects a selective application of inclusion logic. This inconsistency may inadvertently exaggerate the CI effect by misrepresenting the true cognitive demands associated with high-interference conditions.

Taken together, such decisions—unless explicitly acknowledged and justified—undermine the very definitional precision that Czyż, (2025) advocate and raise question about the consistency of their methodological critique.

Classification of Serial Practice Within the Broader High Contextual Interference Framework and Associated Explanatory Mechanisms Considerations

The commentary raised concerns about our inclusion of serial practice conditions within the broader "high CI" category, suggesting that serial and random schedules differ in their underlying mechanisms and therefore should not be grouped analytically. While this conceptual distinction is acknowledged, we provide several arguments supporting the theoretical basis for our approach and respectfully disagree that it compromises the integrity or interpretive value of our meta-analysis.

First and foremost, we acknowledge—and have previously stated—that serial practice lies conceptually between blocked and random formats. Specifically, as outlined in Ammar et al. (2023): "The earliest experiments in CI distinguished between two CI levels… A serial practice group was added to the CI paradigms by Lee and Magill (1983), wherein participants perform, during the same acquisition session, several tasks in serial order (e.g., AAAAAAAAA or ABCABCABC)." We are thus fully aware of its intermediate status and clearly communicate this to readers, including the nuanced theoretical debate surrounding its classification.

Furthermore, we recognize that serial and random practices may rely on partially distinct cognitive mechanisms—a distinction well documented in the early motor learning literature (e.g., Klapp, 1995, 2003). However, as our analysis focused on behavioral outcomes—namely acquisition, retention, and transfer—rather than isolated neurocognitive processes, we adopted a broader CI framework in line with theoretical and empirical precedent. There was no suggestion in our paper that serial practice is equivalent to random practice. Instead, we followed a well-established conceptual approach in the motor learning literature: both serial and random formats are recognized as inducing greater contextual interference than blocked practice. Accordingly, we classified them under a shared "high CI" umbrella for analytical comparison against "low CI" (i.e., blocked) conditions. This broader classification is supported by foundational theoretical work (e.g., Lee & Magill, 1983) and later meta-analytical syntheses (e.g., Lage et al., 2021), which suggest that both random and serial schedules produce greater retention and transfer gains than blocked conditions—even if the underlying mechanisms (e.g., elaboration vs. predictability) differ.

From the outset of our manuscript, we explicitly described this theoretical background. Furthermore, we clearly outlined the potential explanatory mechanisms for the CI effect. The mechanisms behind the well-established theoretical position that "practice under blocked conditions facilitates the acquisition of skilled performance and impairs retention and transfer; however, practicing under random or serial schedules leads to inferior acquisition, but superior retention and transfer capabilities" (Battig, 1966; Lee & Magill, 1985; Schmidt et al., 2011; Shea & Morgan, 1979) were introduced early in the "Introduction" section. These theoretical and potential mechanisms-based explanations were not only stated but also elaborated in the Discussion section to help interpret and explain our findings. There, we discussed both supportive and divergent results relative to the CI hypothesis, referencing cognitive mechanisms such as elaboration and reconstruction, as well as task-specific moderators. This makes the critique of our work for not elaborating on mechanisms appear inconsistent—especially since our approach was explicitly grounded in, and discussed within, the broader CI framework.

By contrast, the methodological stance taken in the reviews by Czyż et al., (2024a, b) appears selectively restrictive. They themselves declared a strict focus on blocked vs. random practice, choosing to exclude serial conditions—even from studies already included in their dataset. This decision appears to reflect a non-comprehensive approach that contradicts their own emphasis on mechanisms differentiation, as claimed in their current commentary on our work. If mechanism-related concerns were central to their interpretation, it would have been both appropriate and valuable to include and compare serial, random, and blocked practice conditions side by side. Counterintuitively, their exclusion of serial formats, limited the very underlying mechanisms exploration they now suggest should have been pursued. Moreover, although their reviews included several laboratory-based studies, the discussion in both papers was largely limited to summarizing effect sizes and comparing their results to prior meta-analyses (including ours), with minimal elaboration on theoretical or mechanisms implications. This lack of integration between cognitive theory and empirical synthesis further underscores questions about the foundation of their critique.

In summary, our classification of serial and random practice under a unified "high CI" category was not arbitrary but methodologically justified, transparently reported, and grounded in decades of CI literature. We emphasize again that our analytic goal was to contrast higher vs. lower levels of contextual interference—not to imply that all high CI schedules operate through identical mechanisms. Our interpretations accounted for this heterogeneity and reflected the ecological diversity of real-world learning environments, while maintaining our primary objective: to examine the behavioral learning outcomes associated with high vs. low CI schedules. Rather than undermining our conclusions, this inclusive and ecologically sensitive approach broadens the theoretical relevance and practical utility of our findings.

Definition of the Analyzed Context and Justification for Study Inclusion

The concern raised about a potential mismatch between our abstract statement referring to "physical education (PE) and sports contexts" and the broader inclusion of studies involving gross motor skills warrants clarification. Our inclusion criteria—clearly outlined on page 7 of the manuscript—specified that any outcome evaluating the acquisition, retention, or transfer of learned sports/gross motor skills was eligible. Gross motor skills were defined, as per Esposito and Vivanti (2013), as movements engaging large muscle groups (e.g., balance, walking, jumping, throwing), which underpin a wide range of physical tasks, including those in sport and PE.

Accordingly, Pollatou et al. (1997) and Jeon et al. (2021), despite being framed as "novel" or "balance game" tasks, both involved performance of well-established gross motor functions: throwing/kicking (Pollatou et al., 1997) and dynamic balance/mobility movements (Jeon et al., 2021). These clearly align with our operational definition of gross motor skills. Such skills are central to physical literacy and are frequently targeted in educational and training programs, including in adapted PE or geriatric training environments.

As for Krause et al. (2014), although the intervention manipulated feedback rather than the sequence of practice tasks, the feedback manipulation itself was structured in a blocked vs. serial format. The authors explicitly framed their experimental manipulation as a CI intervention and compared its effects across novice and expert golf putting performance (e.g., distance and angle)—a well-recognized sport skill. According to the reprogramming explanation hypothesis (Lee & Magill, 1983), increased feedback variation can be assumed to result in greater utilization of working memory, leading to an interference effect, which is, in turn, posited as the cause of a subsequent enhancement in the learning effect. The theoretical framing provided in their introduction—citing foundational CI studies (Lee & Carnahan, 1990; Lee & Magill, 1983; Shea & Morgan, 1979)—and the nature of the motor task (golf putting) clearly position the study within the broader CI and motor learning literature. Furthermore, Krause et al. explicitly hypothesized differential effects of feedback structure on retention, a core outcome in CI paradigms.

While the manipulation of feedback might be seen as an indirect influence on motor performance, similar to error processing, it was nonetheless designed to impact motor learning through varying cognitive load on working memory, rather than purely affecting perceptual or declarative knowledge (Broadbent et al., 2017). This approach aligns with our inclusion criterion of interventions examining the effect of practice structure on performance in a gross motor skill, provided the measured outcome is motor-based.

In this light, all three studies were methodologically and conceptually appropriate for inclusion. The context in which gross motor skills are practiced—whether in physical education programs, or sports training—is not contradictory but rather complementary, as gross motor skills form the foundational basis upon which both physical education and sport-specific competencies are built. Therefore, our inclusive framing captures the full spectrum of environments where such learning occurs and reflects the shared motor learning principles underlying all these domains.

Thus, we respectfully argue that the definition of the analyzed context was not blurred but intentionally broadened to capture ecologically and educationally relevant gross motor learning scenarios across settings. Excluding such studies would have unjustifiably narrowed the evidence base and limited the generalizability of our conclusions regarding CI effects on motor performance.

Notably, a convincing counterexample to the criticism by Czyż (2025) regarding the "blurred definition of the analyzed context, including inconsistency between study objective and inclusion criteria" in our review can be drawn from their own blurred understanding of the term "general population," which reveals an inconsistency between their stated objective and their applied inclusion strategy. Indeed, in the rationale of their meta-analysis, Czyż et al. (2024a) emphasize the importance of examining the CI effect specifically in the general population, stating: "it is crucial to update and perform a meta-analysis focusing on the CI effect on retention in the general population…" Beyond the epistemologically problematic nature of such a broad claim—particularly in light of statistical generalizability concerns (Acree, 2021)—this language suggests an intention to generate findings with wide applicability, presumably inclusive of clinical groups, older adults, and individuals with varying health statuses. This is especially relevant given the increasing interest in CI effects in contexts such as rehabilitation, neuroplasticity, and motor relearning.

However, in contrast to this broad objective, their inclusion criteria sharply restrict the analysis to "only healthy participants, as classified by the authors of the primary studies." This introduces a definitional and methodological inconsistency. In the established scientific literature, the term general population is typically understood to include both healthy individuals and those with medical conditions or risk factors. For instance, Rothman et al. (2008) in Modern Epidemiology describe the general population as including individuals "irrespective of risk status or diagnosis." This inclusive understanding aligns with policy guidance from the US National Institutes of Health (NIH), which mandates the inclusion of individuals across the lifespan including minorities groups, particularly in clinical research, unless their exclusion is scientifically justified (NIH, 2001). Similarly, the CONSORT 2010 guidelines emphasize the need for transparent reporting of participant characteristics to enhance the applicability and generalizability of study findings (Schulz et al., 2010). These frameworks advocate for the inclusion of diverse subgroups in research to ensure that findings are broadly applicable and support equitable scientific conclusions.

By excluding all clinical or at-risk populations despite claiming to address the general population, Czyż et al. (2024a) significantly narrow the scope and generalizability of their findings. This creates another inconsistency in their systematic review strategy, especially when juxtaposed with their critique of our work for not explicitly delimiting PE and sport contexts—which, in contrast, are logically embedded within the broader gross motor skills domain we studied. Their own "blurred population definition" undermines the strength of their methodological critique and highlights the importance of aligning eligibility criteria with the declared aim of a systematic review.

Interpretation of Post-Test Results Within the CI Framework

The critique concerning our inclusion and interpretation of post-test results reflects a misunderstanding of both the objectives and methodologies used in our meta-analyses (Ammar et al., 2023, 2024). It is worth noting that this point again extends beyond the intended scope of their commentary on our 2024 publication and targets our earlier work from 2023—which, while inappropriate in context, is also inaccurately characterized.

Czyż (2025) suggests that using post-test results to evaluate the CI effect is "misleading," citing sleep-dependent memory consolidation as a justification for focusing exclusively on delayed retention assessments. While we agree that sleep and consolidation processes are central to long-term motor learning, their comments oversimplify our analytical framework and overlook key aspects of our methodological design and theoretical grounding.

First, both Ammar et al. (2023) and Ammar et al. (2024) were designed to assess the CI hypothesis comprehensively—across all critical time points: post-test (acquisition), retention (relatively permanent gains), and transfer (generalizability). This structure aligns precisely with the established theoretical foundations of CI, which posit that low CI (blocked practice) facilitates superior acquisition, while high CI (random/serial practice) enhances retention and transfer (Battig, 1966; Lee & Magill, 1985; Schmidt et al., 2011; Shea & Morgan, 1979).

In the 2023 meta-analysis, Ammar et al. analyzed performance at post-test, retention, and transfer across eligible studies with careful attention to baseline equivalence, thereby ensuring valid group comparisons. We did not, as implied by Czyż (2025), treat post-test outcomes in isolation as indicators of "learning" or long-term gains. Rather, post-test scores were analyzed as measures of acquisition performance, explicitly separated from longer-term measures of retention and transfer, which were also included (Ammar et al., 2023).

In the 2024 meta-analysis, currently under critique, we extended this evaluation by computing change scores across multiple phases: pre- to post-test (acute acquisition gain), post- to retention (sustainability of learned skills), and pre- to retention (total learning gain). This approach allowed us to assess not only whether high CI improved delayed retention, but also the degree to which learning gains were preserved or deteriorated across time—a critical and underexplored element in previous reviews. The rationale for this three-phase structure was clearly defined and justified in the Introduction and Methods sections of our paper (Ammar et al., 2024).

Counterintuitively, while Czyż (2025) highlight the need for comprehensive and temporally informed analysis of CI effects, their own reviews fall short of this standard. In their 2024 meta-analyses, Czyż et al. focused exclusively on retention (in one study; Czyż et al., 2024a) and transfer (in the other; Czyż et al., 2024b), without including the acquisition phase or reporting any analysis of pre-to-post or post-to-retention change scores. Moreover, their work did not ensure or control for baseline equivalence between random and blocked groups—an omission that could affect their conclusions, as detailed in point 1 of the Point-by-Point Rebuttal section.

In contrast, Ammar et al. (2024) explicitly acknowledged the importance of considering both short-term and long-term changes in performance, as consistent with Schmidt and Lee's (2019) definition of motor learning: "a set of processes associated with practice or experience leading to relatively permanent gains in the capability for skilled performance." It follows logically from this definition that analyzing both acquisition and retention phases—and the change between them—is essential to a more complete understanding of learning efficacy under different practice conditions.

The critique also highlights another fundamental inconsistency in Czyż et al., (2024a, b)'s own conclusions. Despite aiming to evaluate CI "in the general population," they draw sweeping generalizations from data collected at only one time point (retention or transfer), while failing to address confounding variables such as group equivalence or sleep timing. Furthermore, while criticizing our use of post-test data, they do not explain why their reviews omit acquisition assessments entirely—despite acquisition being a fundamental component of the CI model. Importantly, the term contextual interference was not arbitrarily chosen; it was explicitly intended to capture the essential phenomena of the CI approach (Battig, 1979). As originally conceptualized by Battig (1966, 1972, 1979), learning occurs when the acquisition of a task (the text) is disrupted by other tasks (the context), producing interference during and after acquisition that ultimately enhances subsequent learning advantages. This foundational mechanism—interference during and after acquisition—is not ancillary, but necessary for the beneficial learning outcomes associated with CI. How can one claim to understand or comprehensively evaluate the CI approach while disregarding the very phenomenon—interference during and after "acquisition"—that defines its core mechanism? (Schöllhorn, 2016; Schöllhorn et al., 2022).

In summary, (i) we did not equate post-test scores with "learning benefits," but treated them as indicators of acquisition—separate from retention and transfer measures; (ii) both Ammar et al. meta-analyses evaluated all critical time points to test the CI hypothesis in a holistic and theory-driven way; (iii) our 2024 review extended the field by analyzing performance changes over time (pre-post, post-retention, and pre-retention), offering more nuanced insights into learning sustainability; (iv) in contrast, Czyż et al., (2024a, b)'s reviews neglected acquisition entirely and did not account for key confounds, thereby limiting their scope and interpretive power; and (v) while Czyż et al. critique the breadth and comprehensiveness of our approach without sufficient evidence, their own methodology does not fully reflect the standards they advocated.

Therefore, we respectfully disagree with the claim that our use of post-test data could affect the validity of our findings. On the contrary, our inclusion of all three critical learning phases strengthens the conceptual and practical relevance of our work.

Brief Comparative Evaluation of Recent CI Related Meta-analysis and Reflection on Meta-analytics Practice in Motor Learning

Recent years have witnessed a surge in meta-analytic studies exploring the CI effect on motor learning. Among the most prominent are the works of Ammar et al., (2023, 2024) and Czyż et al., (2024a, b), which, despite addressing similar research questions, arrive at notably different conclusions. This divergence can be largely attributed to differences in their methodological scope, theoretical orientation, outcome operationalization, and statistical strategies. While Ammar et al., (2023, 2024) sought to evaluate the generalizability of the two CI effects in gross motor skills—primarily in applied settings such as PE and sport—Czyż et al. focused the second aspect of the CI hypothesis (i.e., retention and transfer) across laboratory and applied settings.

To better understand the basis for these differences, we present a detailed comparative table (see Table 1) that outlines key distinctions across four recent meta-analyses in terms of focus, outcomes assessed, statistical modeling, treatment of confounding factors (e.g., baseline equivalence), and conclusions drawn. The table also details how each study approached the inclusion of relevant subgroups, quality control, and interpretation of moderator effects.

Scope, Strengths, and Limitations

A comparative evaluation of four recent meta-analyses on CI effect highlights distinct rational, methodological approaches, analytical scopes, and interpretive frameworks. Ammar et al., (2023, 2024) offer strong, theory-informed, and methodologically novel justifications for their reviews. In contrast, Czyż et al., (2024a, b) lack originality and are positioned more as critiques of previous work rather than as independent, necessary syntheses. Ammar et al. (2023) delivered a comprehensive synthesis of CI effects across the three fundamental phases of motor learning: acquisition, retention, and transfer. Their study emphasized gross motor skills performed in applied settings, particularly PE and sports contexts, and focused on outcome-specific timepoints. A major methodological strength was their exclusion of studies lacking baseline equivalence between intervention and control groups, which helped reduce potential confounding effects. Furthermore, they incorporated across a relatively large sample size (>1100 participants), a range of potential moderators including participant characteristics (i.e., age, sex, and experience), acquisition protocol (i.e., practice duration, number of training sessions and acquisition trials, duration between testing sessions, and testing protocol), and sports/skills being investigated. However, the study did not employ multilevel meta-analysis and treated multiple outcomes from a single study as independent, a limitation that may have introduced statistical dependencies.

Ammar et al. (2024) extended this work by analyzing not only absolute scores but also skill performance gains using change scores across different learning phases (pre-post, post-retention, and pre-retention). They prioritized practical relevance for

Table 1. Overview of Ammar et al. and Czyż et al. systematic review and meta-analyses on CI effects

Aspect	Ammar et al. (2023)	Ammar et al. (2024a)	Czyż et al. (2024a)	Czyż et al. (2024b)
Rationale and novelty	Novel and clearly articulated rationale: The first meta-analysis focused explicitly on the effects of CI in gross motor skills, directly addressing the limitations of earlier reviews (e.g., Brady, 2004); The paper critiques prior syntheses for their lack of task-domain differentiation and introduces updated coding procedures and outcome criteria.; The contribution is timely, field-specific, and advances the literature by clarifying the domain of application and offering a better-grounded synthesis.	Building on the 2023 review, this study provides a methodologically and conceptually advanced synthesis. It introduces multilevel meta-analytic modeling with a key novelty by focusing on performance gains (delta scores) rather than isolated time-point measures, allowing more accurate tracking of learning effects. The paper positions itself as a transparent and theory-driven analysis, responding directly to methodological gaps in previous CI research with broader contextual and practical considerations.	The stated rationale revolves around updating prior syntheses (e.g., Brady, 2004) in light of accumulated studies. However, the novelty is limited, as the review covers similar ground to Ammar et al., 2023, which was published shortly beforehand. The authors frame their review as necessary due to alleged methodological flaws in Ammar et al., particularly regarding database selection and preregistration. However, as clarified in part A of the present paper, this critique is misleading, as no valid registration standard was violated and multiple databases were searched. Thus, the rationale for the new review appears more reactive than driven by a gap in the literature.	Focusing on transfer outcomes, closely follows the 2024a paper and again lacks novelty, particularly as Ammar et al. (2023, 2024) had already addressed both retention and transfer effects. The justification is again framed around perceived shortcomings in Ammar et al.'s reviews rather than a clearly defined conceptual gap. This results in a rationale that appears retrospective and confirmatory, with minimal advancement in scope or method. No new conceptual or methodological tools are introduced.
Primary focus	CI effect generalizability in gross motor skill (e.g., physical education and sports contexts). Analyzes absolute performance outcome at post-test, retention-test and transfer-test.	CI effect generalizability in gross motor skill (e.g., physical education and sports contexts). Analyzes acute and relatively permanent performance gains.	CI effect on delayed retention in motor learning in general. Aims to update Brady et al. (2004) and solve controversies.	CI effect on delayed transfer in motor learning in general. Focuses exclusively on transfer.
Outcomes analyzed	- Acquisition - Retention - And/or transfer performances	- Δ pre-post - Δ post-retention - Δ pre-retention tests. (Δ): Performance gains	- Delayed retention test performance (tests after 24 h). (Immediate retention discussed in systematic review but not meta-analysis.)	- Delayed transfer test performance (tests after 24 h). (Immediate transfer discussed in systematic review but not meta-analysis.)
Overall conclusion	- CI effects in sports are inconsistent and context-dependent - Offering no clear overall advantage for CI in acquisition, retention, or transfer.	- Very limited evidence supporting high CI for longer-term performance advantage. - Question the generalization of the CI model to PE and sports. - No statistically	Concluded high CI has a medium beneficial effect on delayed retention for the whole population (including all age groups in both applied and laboratory settings), which was statistically significant. Describes the CI effect as robust based on	Found a medium, statistically significant overall CI effect on delayed transfer favoring random practice (SMD = 0.55).

applied settings and adopted a multilevel meta-analytic approach, accounting for hierarchical data structure and correcting for effect directionality and outliers. The study also maintained the exclusion of studies with baseline inequality, enhancing internal validity. Similarly, they incorporated across a relatively large sample size, a

Table 1. (continued)

		significant overall difference for Δ pre-post or Δ pre-retention. - Significant difference favoring high CI for Δ post-retention, but not large enough for overall greater long-term gain (Δ pre-retention).	delayed retention.	
Findings: applied settings	Only around 20% of outcomes supported a CI advantage: 18 of 103 acquisition measures, 19 of 84 retention measures, and 6 of 36 transfer measures. Small, non-significant effects of random versus blocked practice were observed in applied sports settings for acquisition (effect size [ES] = 0.10), retention (ES = − 0.16), and transfer (ES = − 0.24), with most outcomes showing no difference.	- Very limited evidence supporting CI generalization for acute or relatively permanent gains in PE/sports settings. - Only 20% of outcomes agreed with paradoxical CI effects in PE/sports.	Beneficial effect of random practice on retention in the applied setting was almost negligible and statistically non-significant (SMD = 0.35 or 0.30). Applied settings yielded fewer convincing results than laboratory studies.	ES for transfer in applied studies was small and statistically non-significant (SMD = 0.34). Less significant effects in applied settings potentially due to task complexity or other factors.
Findings: laboratory settings	The focus was on gross motor skills, mainly relevant to applied setting.	The focus was on gross motor skills, mainly relevant to applied setting.	Random practice was favored for retention, with a large, statistically significant effect size.	Random practice was favored for transfer, with a medium, statistically significant effect size (SMD = 0.75). Larger effect sizes in laboratory settings potentially due to better control and simpler tasks.
Findings: age group	Age did not overall moderate acquisition (p = 0.232) or retention (p = 0.820) effects. However, 20–24-year-olds showed CI effects with blocked practice enhanced acquisition (ES = 0.282, p = 0.030) while random practice enhanced retention (ES = − 0.405, p = 0.011). Retention gains from random practice also appeared in 25–32-year-olds (ES = − 0.605, p = 0.013). In contrast, adults > 65 consistently favored blocked practice at both acquisition (ES = 1.356, p < 0.001) and retention (ES = 0.943, p = 0.008).	Age significantly moderated both acquisition (Δ pre-post) and permanent gain (Δ pre-retention) effects (p < 0.0001 for both). Limited support for paradoxical CI effects in specific adult age ranges (20–24, 25–32) for acquisition or permanent gains; paradoxical effects not found concurrently for both gains in any subgroup. Suggested blocked practice favored in older adults in some conditions.	Found significant CI effects for retention favoring random practice in adults (medium SMD) and older adults (large SMD = 1.45). Effect for young participants was negligible and statistically insignificant. Noted contradiction with Ammar et al.'s findings for older adults.	Found significant CI effects for transfer favoring random practice in adults (medium SMD = 0.54) and older adults (large SMD = 1.28). ES for young participants was negligible (SMD = 0.12).

range of potential moderators related to participant characteristics, acquisition protocol, and sports/skills. Although not preregistered, Ammar et al. (2024) followed a transparent, pre-specified protocol. Similar to the 2023 review, the study did not include fine motor skills and missed a very few studies; however, those omissions were judged to have no impact on conclusions.

Table 1. (continued)

Findings: experience level	Experience level did not moderate acquisition (p = 0.731). Yet, inexperienced learners achieved significantly better acquisition with blocked practice (ES = 0.186, p = 0.029). No retention or transfer effects emerged.	Did not influence CI effects in either pre–post or pre–retention phases.	☒ Did not account for experience level	☒ Did not account for experience level
Findings: sports/discipline	Sport type did not moderate acquisition (p = 0.455), retention (p = 0.288), or transfer (p = 0.344). Nonetheless, baseball players performed better post-acquisition following random practice (ES = − 0.917, p = 0.006), and golfers retained skills better after random practice (ES = − 0.771, p = 0.014). The only transfer advantage was for serve skills under random practice (ES = − 0.556, p = 0.010).	Only specific sports showed divergent CI effects in pre–post gains. In kneeling/pistol shooting, blocked practice outperformed random (ES = 0.51, p = 0.017), likely reflecting the precision and consistency demands of shooting. By contrast, baseball swing skills improved more under random practice (ES = − 1.11, p = 0.005). No sports moderated pre-retention outcomes.	☒ Did not account for sports/discipline	☒ Did not account for sports/discipline
Findings: skill type	Skill category did not moderate acquisition (p = 0.426), retention (p = 0.482), or transfer (p = 0.322). Yet "other" skills (balance, cartwheel, etc.) favored blocked practice at acquisition (ES = 0.691, p = 0.006) and random practice at retention (ES = 0.830, p = 0.011). Golf putting/chipping (ES = – 0.766, p = 0.011) and kicking (ES = − 1.248, p = 0.025) also showed retention gains from random practice. Serve skills alone showed a transfer gain.	Did not influence CI effects in either pre–post or pre-retention phases.	☒ Did not account for skill type	☒ Did not account for skill type
Findings: testing protocol	Testing protocol did not moderate retention (p = 0.250) or transfer (p = 0.074). However, retention was significantly better following random practice when assessed with a purely random (ES = –0.451, p = 0.039) or block/random (ES = − 0.479, p = 0.042) protocol.	Did not influence CI effects in either pre–post or pre-retention phases.	☒ Did not account for testing protocol	☒ Did not account for testing protocol

In contrast to both studies of Ammar et al., (2023, 2024), Czyż et al. (2024a) focused solely on delayed retention performance, thereby testing only part of the CI hypothesis. While the study was preregistered and considered age and setting as moderators, it did not verify or statistically adjust for baseline equivalence between groups. This omission potentially may have over-estimated CI effect estimates. Moreover, their review excluded or failed to retrieve several relevant applied-setting

Table 1. (continued)

Findings: sex (e.g., female proportion)	Meta-regression showed no significant moderation by the proportion of female participants at any phase: acquisition (coefficient = 0.005, $p = .068$), retention (coefficient = 0.005, $p = .116$), or transfer (coefficient = 0.006, $p = .243$).	Only the proportion of female participants moderated Δ pre-post gains (coef = 0.0084, SE = 0.003; 95% CI [0.002, 0.015]; t(98)=2.64, $p = .0095$), with higher female representation linked to relatively better low-CI gains. No effect on Δ pre-retention ($p > .05$).	☒ Did not account for sex	☒ Did not account for sex
Findings: number of acquisition sessions	Session count was not a significant moderator in acquisition (coefficient = – 0.009, $p = .506$), retention (coefficient = 0.0385, $p = .107$), or transfer (coefficient = − 0.003, $p = .944$).	Did not influence CI effects in either pre–post or pre-retention phases.	☒ Did not account for session numbers	☒ Did not account for session numbers
Findings: intervention duration	The length of the practice intervention did not influence outcomes: acquisition (coefficient = – 0.003, $p = .510$), retention (coefficient = 0.011, $p = .147$), or transfer (coefficient = − 0.003, $p = .759$).	Total training duration did not moderate Δ pre-post or Δ pre-retention outcomes (all $p > .05$; ES < 0.10).	☒ Did not account for intervention duration	☒ Did not account for intervention duration
Findings: total acquisition trials	The total number of trials practiced did not affect practice scheduling effects: acquisition (coefficient = 0.000, $p = .899$), retention (coefficient = 0.000, $p = .420$), or transfer (coefficient = − 0.001, $p = .107$).	The total number of practice trials did not moderate CI effects on Δ pre-post or Δ pre-retention performance (all $p > .05$; ES < 0.10).	☒ Did not account for acquisition	☒ Did not account for acquisition
Findings: duration between pre- and post-test sessions	The time interval between the pre- and post-acquisition tests failed to moderate outcomes at acquisition (coefficient = − 0.0029, $p = .440$), retention (coefficient = 0.012, $p = .264$), or transfer (coefficient = 0.041, $p = .287$).	The time between pre- and post-tests did not moderate CI effects in Δ pre-post or Δ pre-retention phases (all $p > .05$; ES < 0.10).	☒ Did not account for duration between pre- and post-test sessions	☒ Did not account for duration between pre- and post-test sessions
Databases searched	Scopus, Web of Science, PubMed, Taylor and Francis, SciELO, with additional manual research	Scopus, Web of Science, PubMed, Taylor and Francis, SciELO, with additional manual research	Scopus, Web of Science, EBSCO, ScienceDirect, Google Scholar, PsycINFO, with additional manual search.	Scopus, Web of Science, EBSCO, ScienceDirect, Google Scholar, with additional manual search.
Handling dependent ES	Used a random-effects model.	Used a three-level meta-analytic model. Assumed $r = 0.50$ for calculating	Applied a three-level mixed model and random-effects model with averaged effect	Used a three-level meta-analytic model.

studies, most often without clear justification. Despite these limitations, the authors concluded in general that random CI has a robust and significant effect on retention, a claim that appears overly generalized given the weak and only partially tested CI effects found in applied subgroups.

Table 1. (continued)

		SMD for gain scores.	sizes. Explicitly used a three-level model for primary analyses. Results from both models were very similar.	
Protocol registration	☒ Not mentioned in the source.	☒ Not mentioned in the source.	Study was preregistered.	Study was registered in PROSPERO (CRD42021228267).
Handling serial practice	Categorized serial practice under the broader high CI group for comparison against low CI.	Categorized serial practice under the broader high CI group for comparison against low CI.	Excluded studies describing serial practice order instead of random schedule.	Excluded studies describing serial practice order instead of random schedule.
Handling baseline equivalence	Exclude studies failed to guarantee equivalent baseline between low and high CI groups	Exclude studies failed to guarantee equivalent baseline between low and high CI groups	Study which failed to guarantee equivalent baseline were included, with no relevant statistical analysis (e.g., stratify subgroups based on reported pre-test comparability)	Study which failed to guarantee equivalent baseline were included, with no relevant statistical analysis (e.g., stratify subgroups based on reported pre-test comparability)
Number of studies (quantitative synthesis)	37 studies included in quantitative synthesis yield to total of 225 overall pooled outcomes. Used 103 ES for post-test. Used 84 ES for retention-test. Used 38 ES for transfer-test.	35 studies included in quantitative synthesis (out of 36 in review, one outlier excluded). Used 26 studies yielding 76 ES for Δ post-retention after outlier removal. Used 35 studies for Δ pre-post (implied by Table 2 and abstract). Total of 183 overall pooled outcomes.	54 articles included in quantitative analysis (out of 59 in review). Yielded 194 effect sizes for overall delayed retention analysis. Of which 24 articles assessing applied setting yielded 105 ES	34 studies included in quantitative analysis (out of 42 in review). Yielded 86 effect sizes for overall delayed transfer analysis. Of which 17 articles assessing applied setting yielded 40 ES
Discussion of theoretical or mechanisms-based explanations of findings	Discusses neurocognitive and learning mechanisms to explain variation in CI effects, including task complexity, attentional control, and contextual cues. Offers some interpretation grounded in cognitive load theory and the interaction between variability and learning stages.	Provides a comprehensive theoretical interpretation that goes beyond surface-level statistical findings. Integrates schema theory, elaboration and reconstruction hypotheses, and practice specificity, offering a multi-layered account of observed effects. Also discusses potential moderating mechanisms (e.g., skill level, feedback). This discussion is deep, theory-driven, and relevant for application.	Provides limited theoretical elaboration, focusing on effect sizes and subgroup results (e.g., age groups). Some mention of neural activation differences and CNS involvement, but these are not fully developed or connected to the meta-analytic findings. Explanatory mechanisms and Interpretation remains general and descriptive.	Slightly more explanatory than 2024a but still lacks depth. The discussion does not systematically relate findings to underlying mechanisms or learning theories. Mentions prior models (e.g., Battig) but seem to fail to connect data trends to cognitive or neural mechanisms in a structured way.

Similarly, Czyż et al. (2024b) focused exclusively on transfer effects, offering yet another partial examination of the second CI hypothesis based on an alternative operationalization of learning—one that must be clearly distinguished from the first CI effects on acquisition-related gains. This review, though preregistered, shared similar shortcomings: lack of baseline control, limited moderator analysis

Table 1. (continued)

Methodological considerations	Clearly includes a well-structured "Methodological Considerations" subsection. It outlines limitations such as possible publication bias, sample representativeness, and task heterogeneity. Importantly, it reflects critically on the impact of moderators and the generalizability of findings, providing actionable insight for interpretation. The discussion is methodologically grounded and appropriately cautious, helping to interpret effect size heterogeneity.	The methodological considerations are not labeled as a separate subsection but are comprehensively embedded across the discussion and conclusion. The paper discusses the rationale for focusing on gain scores, addresses limitations of effect size interpretation, and clarifies the strengths of using multilevel modeling. The explanation of model assumptions and their implications is technically sound, demonstrating strong methodological awareness.	The paper includes paragraphs that seems to address methodological issues but lacks a clearly titled or deeply developed "Methodological Considerations" section. There is minimal reflection on the generalizability, risk of bias, or assumptions behind the analytical approach.	Similarly lacks a dedicated or detailed methodological reflection. Although it emphasizes the claim of improved methods over prior meta-analyses, this assertion is not well substantiated by in-depth analysis or critical evaluation. There is little technical detail on how their methods improve reliability or validity beyond general statements.
Practical application	Includes a clearly identified section on practical implications. The authors relate findings to real-world motor learning and training settings (e.g., sports and rehabilitation), discussing when and how to apply CI based on task characteristics. This adds value for practitioners and researchers, translating findings into actionable advice.	Provides strong practical reflections embedded in the discussion, including how the use of gain scores offers better insight for applied interventions. It also includes recommendations for coaches and therapists about practice conditions that enhance skill acquisition. Although not formatted as a separate section, the practical messages are clear and directly linked to statistical results.	Provides minimal practical application insight. Focus is primarily on methodological critiques of previous work. There is little to no engagement with how their findings should inform practice.	Slightly better than 2024a in that it discusses transfer implications briefly, but again lacks a distinct, well-developed practical application section. The advice to practitioners remains vague and not operationalized.
Future perspective	Concludes with a concise but impactful call for future research directions, grounded in both theoretical advancement and methodological refinement. Two major contributions are particularly notable: (1) the recommendation to employ neurophysiological measures such as EEG in dynamic sport-specific contexts to better understand the neural substrates of CI, and (2) an explicit call to account for individual differences in responsiveness to CI protocols, highlighting a precision-learning approach. These future directions go beyond standard calls for replication and offer a clearly articulated research	Builds upon the 2023 study with an even more detailed and theory-informed roadmap for future research. The authors reinforce the need for EEG-based investigation in ecologically valid settings, underscoring the importance of moving beyond behavioral metrics toward mechanisms understanding of CI effects. They also stress the significance of modeling individual variability, advocating for more personalized motor learning interventions and analytically advanced	Offers a brief and generic paragraph mentioning the need for more research with stronger methodological rigor, but lacks concrete, operational suggestions. No mention is made of neurophysiological tools, individual variability, or advanced modeling. The emphasis remains on critiquing earlier work rather than laying out a forward-looking agenda. This results in a missed opportunity to build constructively on their findings.	Provides only minimal guidance for future research. Although it recommends improving methodological quality, it does not specify concrete tools, target populations, or innovative methodologies that could meaningfully advance the field. There is no mention of potential explanatory mechanisms or neurophysiological approaches, nor of adapting interventions to individual differences. Overall, its future outlook is vague and lacks depth. Moreover, the suggestion to focus future research on young participants in laboratory settings and older participants in applied

(focused only on age and setting), and incomplete transparency in study exclusion decisions. Thus, while both Czyż et al. reviews adopted advanced three-level

Table 1. (continued)

	agenda with applied and mechanisms-related relevance. This perspective aligns with emerging trends in motor learning and neuroscience and suggests integrative experimental designs that are still largely unexplored in the CI domain.	designs (e.g., hierarchical/multilevel models). These recommendations are not only technically sound but also strategically aligned with the push for transdisciplinary approaches in cognitive-motor research. The future perspective here is ambitious yet realistic, calling for designs that can validate the neural and cognitive underpinnings of CI effects across populations.		settings presents an unclear and unjustified limited environmental distinction that lacks a solid theoretical or empirical rationale

meta-analytic models, their narrower CI effects' scope and methodological limitations call for cautious interpretation of their generalized conclusions.

In terms of discussion content and methodological consideration, Ammar et al., (2023, 2024) offer a much stronger and integrated theoretical discussion, aligning empirical results with well-established motor learning models while also offering solid methodological insight. Czyż et al., (2024a, b) papers are more descriptive and empirical, with minimal theoretical synthesis and methodological reflection. Similarly, considering offered practical applications and future directions, Ammar et al., (2023, 2024) papers clearly aim to bridge theory and practice and offer clear, specific, and well-reasoned future directions, helping shape the research agenda, while Czyż et al., (2024a, b) works do not go beyond broad generalities, limiting their applied impact.

In summary, the Ammar et al., (2023, 2024) reviews offered broader and more ecologically valid assessments of CI effects with greater methodological rigor, particularly in the 2024 review, which applied more advanced meta-analytic techniques. In contrast, the Czyż et al., (2024a, b) reviews, although addressing dependency through a multilevel meta-analysis, were narrower in the range of CI hypotheses evaluated, more constrained in their analytical approach—particularly in terms of baseline controls and moderator analyses—and drew generalizable conclusions that were not entirely consistent with their own subgroup finding.

Key Methodological Differences and Interpretive Implications

Upon close comparison, several methodological distinctions emerge that likely influenced the divergent conclusions drawn by these four meta-analyses:

- Scope of analysis: Ammar et al., (2023, 2024) examined the full CI hypothesis across acquisition, retention, and transfer, whereas Czyż et al., (2024a, b) each focused on a single learning phase, providing only partial validation of CI effects.

- Population and context: Ammar et al., (2023, 2024) centered their analyses on gross motor skills, predominantly within PE and sports settings, ensuring ecological relevance to real-world motor learning. In contrast, Czyż et al., (2024a, b) included both fine and gross motor skills and drew from a mix of laboratory and applied settings. Although they performed adequate analyses to distinguish between these contexts, their overall conclusions were largely based on pooled effect sizes, combining both settings. This approach, while methodologically acceptable, may have obscured context-specific trends—particularly the weaker or null CI effects observed in applied settings—thereby limiting the practical generalizability of their conclusions for educators and practitioners.
- Inclusion of serial practice: Ammar et al., (2023, 2024) included serial practice under the high CI condition based on theoretical grounding from Lee and Magill (1983), while Czyż et al., (2024a, b) excluded these studies, adopting a stricter operational definition of high CI limited to random practice.
- Data inclusion strategies: Ammar et al., (2023, 2024) used tools such as WebPlotDigitizer to recover missing means and standard deviations from graphical data, ensuring broader inclusion. Czyż et al., (2024a, b) excluded such studies, potentially narrowing their datasets and reducing representation of applied research.
- Baseline equivalence: Ammar et al., (2023, 2024) ensured and required baseline equivalence across groups, excluding studies that failed to meet this condition. Czyż et al., (2024a, 2024b), in contrast, included such studies without performing stratification or relevant sensitivity analyses, potentially biasing pooled effect estimates.
- Statistical models: Ammar et al. (2023) used a traditional random-effects model, while Ammar et al. (2024) and both Czyż et al., (2024a, b) reviews adopted advanced three-level meta-analytic models. However, only Ammar et al. (2024) comprehensively addressed effect directionality, outliers, and data dependencies.
- Moderator analysis: Ammar et al., (2023, 2024) considered multiple moderators including age, sex, expertise levels, skill type, acquisition protocols, etc. Czyż et al., (2024a, b) limited their moderator analysis to age and context (applied vs. laboratory), potentially omitting significant moderators' influences.
- Interpretation of CI effects: Ammar et al., (2023, 2024) concluded that CI effects are inconsistent and context-dependent, with limited support for generalization, particularly in applied settings. Conversely, Czyż et al. (2024a) made broader claims of CI robustness despite their own subgroup analyses revealing minimal CI effects in applied settings, indirectly aligning with Ammar et al.'s more cautious conclusions.

Taken together, these methodological choices shaped not only the scope and inclusiveness of each meta-analysis but also their final interpretations. While Czyż et al.'s conclusions assert CI's robustness, their subgroup findings indicate limited support for CI in applied settings. This paradox underscores that the apparent divergence between these reviews stems more from differing methodologies and interpretive frameworks than from fundamentally conflicting evidence.

Reflection on Meta-analytic Practice in Motor Learning

Despite the growing enthusiasm for evidence synthesis in motor learning research, particularly through systematic reviews and meta-analyses, critical reflection on current meta-analytic practices reveals several challenges and concerns. One key issue pertains to the methodological rigor and interpretation of meta-analytic findings in a domain characterized by considerable heterogeneity, small sample sizes, and diverse outcome metrics. These issues are compounded by statistical ritualism and a tendency to misapply or overinterpret conventional tools of bias detection, leading to potentially misleading conclusions.

Recent debates have highlighted the limitations of routine statistical practices in social and behavioral sciences. Gigerenzer (2004, 2018) warned against the "statistical ritual" that fosters replication delusions and could affect genuine statistical reasoning. This critique is particularly relevant for sport science meta-analyses, where over-reliance on *p* values, confidence intervals, and funnel plot asymmetry can obscure rather than clarify the robustness of findings. The visual inspection of funnel plots and application of underpowered statistical tests (e.g., Egger's or Begg's tests) often fail to detect true small-study effects or publication bias, yet they continue to be used indiscriminately even when meta-analyses include fewer than ten studies—a threshold below which such tests are generally unreliable.

In line with these critiques, recent meta-analytic work by McKay et al., (2022a, 2022b) has systematically questioned some of the most widely accepted assumptions in motor learning research—such as the purported benefits of reduced feedback frequency, autonomy support, and self-controlled practice. These authors emphasized the need for more rigorous, well-powered primary studies and urged caution against over-interpreting results potentially shaped by publication bias and methodological flaws. Their overarching message is clear: established theories in motor learning should not be immune to scrutiny, and evidence synthesis must be guided by methodological transparency and critical evaluation.

Afonso et al. (2024) further echoed these concerns by emphasizing the widespread misuse and misinterpretation of publication bias tests in sport and exercise science. They argued for replacing definitive labels such as "publication bias" with more cautious language such as "risk of publication bias," highlighting that funnel plot asymmetry can arise from heterogeneity, outcome metric choice, or random error. Similarly, many meta-analyses apply grading frameworks such as GRADE without sufficient critical scrutiny, potentially downgrading evidence quality based on methodologically limited assessments of bias.

Another concern is the redundancy and fragmentation of meta-analytic work in motor learning. The rapid proliferation of meta-analyses—often addressing overlapping questions without substantially advancing theoretical or practical understanding—can create confusion, inflate perceived evidence strength, and stifle innovative research. Such redundancy often stems from the absence of higher-order evidence syntheses (e.g., umbrella reviews) and the limited incorporation of domain-specific quality standards, such as tailored risk of bias assessments and transparency in data handling.

Moreover, most meta-analyses in motor learning continue to draw from a limited pool of primary studies—many of which are of low methodological quality and exhibit substantial heterogeneity. This includes variability in skill types, participant characteristics, intervention protocols, outcome measures (e.g., error vs. accuracy), and assessment timelines (e.g., immediate post-test vs. delayed retention). Such heterogeneity complicates the synthesis process and increases the risk of inconsistent effect directions and statistical dependencies. Despite these challenges, many existing meta-analyses do not employ multilevel or multivariate modeling strategies to account for nested data structures (e.g., multiple outcomes within a single study). Additionally, corrections for outcome directionality, as well as systematic outlier and influence diagnostics, are often overlooked. Only a few meta-analyses apply robust multilevel models, conduct proper outlier detection, or implement domain-appropriate corrections for interpreting performance scores.

To enhance the rigor and credibility of future meta-analyses—should their execution be justified by the emergence of a greater number of high-quality original studies—several best practices are recommended, spanning both primary research and evidence synthesis.

- Ensure baseline equivalence in original randomized controlled trials (RCTs):

  Group comparability at baseline is crucial to attribute observed differences to the intervention rather than pre-existing performance disparities. Researchers could:

  - Use random and concealed allocation procedures.
  - Report and statistically test pre-test comparability between groups using t tests or ANOVAs.
  - Apply covariate adjustment techniques (e.g., ANCOVA, mixed models) when minor imbalances are detected.
  - Consider change-score analysis (pre-to-post or post-to-retention) to isolate intervention effects independent of baseline levels.

- *Account for baseline imbalance in meta-analyses:*

  Many existing meta-analyses fail to stratify studies based on baseline equivalence or to perform sensitivity analyses excluding imbalanced studies. To address this, researchers should:

  - Stratify subgroups based on reported pre-test comparability.
  - Conduct sensitivity analyses to evaluate the robustness of pooled estimates when potentially biased studies are excluded.
  - Favor inclusion of studies that report within-group change scores or adjusted effect sizes controlling for baseline differences.
  - Incorporate meta-regression models to assess whether baseline imbalance acts as a moderator of the reported effects.

  As highlighted in our recent response to Czyż (2025) in Point-by-Point Rebuttal section, the absence of such techniques "limits confidence in the pooled effect estimates" and may result in overestimated effects of CI.

- Adopt multilevel and multivariate meta-analytic approaches:

Motor learning studies often contribute multiple, non-independent outcomes. Appropriate modeling strategies such as three-level meta-analysis (e.g., Assink & Wibbelink, 2016) or robust variance estimation (RVE) can help account for this dependency. These methods also allow for the modeling of random effects at the study and outcome level, improving estimate accuracy and reducing type I error.

- Perform robust outlier detection and bias correction, most of them already applied by Ammar et al. (2024):

  Identifying influential studies is essential, especially when outliers skew aggregated results. Techniques such as Cook's distance, DFBETAS, or leave-one-out diagnostics should be routinely implemented. For publication bias, beyond funnel plots and Egger's test, researchers could consider:

  - Trim-and-fill procedures.
  - Selection models (e.g., Vevea & Hedges, 1995).
  - Bayesian approaches such as the RoBMA-PSMA framework (McKay et al., 2023, applied).
  - Z-curve analysis to estimate the expected replicability of results and detect selective reporting patterns.

- Develop field-specific quality appraisal tools:

  Common tools such as PEDro or EPHPP do not adequately capture motor learning-specific factors. A tailored tool could include:

  - Assessment of skill type (discrete vs. continuous, open vs. closed).
  - Degree and nature of variability (task variation, noise).
  - Proprioceptive/kinesthetic vs. visual dominance in the task.
  - Participants' experience level and its operational definition.
  - Detail of feedback, instruction, and environmental interaction.

  This would improve the internal validity and ecological relevance of studies included in future meta-analyses.

- *Embrace open science principles and reproducibility standards:*

  - Pre-register protocols (e.g., PROSPERO) with detailed inclusion/exclusion criteria, planned analyses, and data handling decisions.
  - Report deviations transparently, ensuring methodological integrity.
  - Share datasets and coding scripts for peer validation and secondary analyses.

Taken together, these methodological improvements are essential to mitigate the risks of overgeneralization, selective reporting, and statistical artifacts that continue to challenge evidence synthesis in this field. As Gigerenzer (2018) critiques, current statistical "rituals"—including overreliance on p-values and uncritical pooling of inconsistent effects—risk replacing substantive reasoning with mechanical inference, contributing to what he terms the "replication delusion." In motor learning research, where cognitive, perceptual, and biomechanical factors interact in complicated and complex ways, moving beyond surface-level aggregation is not only desirable—it is necessary.

However, meta-analysis is widely used to synthesize findings across studies, yet it replicates many of the problems highlighted by Gigerenzer in his 2004 critique of "mindless statistics." While often seen as the gold standard for summarizing evidence, meta-analyses typically rely on null hypothesis significance testing (NHST), which calculates the probability of data under the assumption that the null hypothesis is true ($P(\text{data} \mid H_0)$). This approach does not provide the probability that the hypothesis itself is true given the data, a misunderstanding that often leads to overconfidence in statistically significant findings. Moreover, meta-analyses tend to increase sample size across studies, thereby amplifying statistical power and reducing *p* values—without necessarily improving the epistemic strength of the underlying conclusions. They often mask heterogeneity, neglect methodological diversity, and ignore deeper theoretical disagreements. This is especially problematic in fields such as motor learning, where studies differ widely in intervention types, task complexity, feedback mechanisms, and time scales.

As an alternative, Feest (2019) suggests a shift away from statistical replication and toward an increase in the number and diversity of arguments for a given hypothesis. This implies valuing heterogeneous methodologies, varied theoretical perspectives, and different experimental conditions—which meta-analysis, by its nature, tends to average out or ignore. For motor learning research, where processes are multifaceted and context-dependent, this pluralistic approach may be more appropriate.

Several frameworks support such an alternative. Bayesian evidence synthesis allows for weighing various sources of evidence and updating belief in a hypothesis, rather than testing against a null (van Wonderen et al., 2024). It accommodates heterogeneity and supports the integration of prior knowledge. Similarly, Inference to the Best Explanation (IBE) emphasizes explanatory power, coherence, and scope over statistical repetition (Lipton, 2000). Robustness reasoning, as proposed by Wimsatt (1981, 2012), argues that the credibility of a hypothesis is strengthened when supported by multiple independent methods that converge on the same result—something particularly relevant in the varied landscape of motor learning studies. Finally, evidential pluralism in the philosophy of science encourages combining experimental, theoretical, observational, and historical evidence to build a more comprehensive understanding (Shan & Williamson, 2022).

In sum, rather than pooling similar studies to increase power, motor learning research would benefit from evaluating a plurality of independent and diverse arguments. Such an approach better acknowledges the inherent complexity of learning processes and enables a richer, more theoretically grounded synthesis. This perspective—more consistent with the enriched, mechanism-focused discussions in Ammar et al., (2023, 2024) than with the narrowly statistical focus of Czyż et al., (2024a, b)—shifts the field away from a reliance on statistical significance and toward a deeper, more meaningful scientific understanding of the studied object of motor learning.

Future Directions in Contextual Interference and the Broad Motor Learning Research

Given the substantial heterogeneity, inconsistent findings, and methodological limitations observed in both primary studies and meta-analyses, the future of CI and broader motor learning research may embrace a more integrated, critical, and theory-driven approach.

First, there is an urgent need for high-quality, well-powered RCTs that explicitly compare different motor learning frameworks (e.g., CI, differential learning, variability of practice) under ecologically valid conditions. These trials should account for baseline equivalence, use repeated measures to track gains across learning phases, and implement reliable outcome assessments that include both quantitative (e.g., accuracy, error scores) and qualitative (e.g., movement quality, neurophysiological correlations) indicators.

Second, the frequent recycling of similar meta-analytic inquiries without addressing deeper theoretical or methodological questions underscores the need to shift towards comprehensive umbrella reviews. Such reviews should synthesize existing meta-analyses, assess consistency across findings, identify overlapping studies, evaluate the certainty of evidence, and highlight unresolved issues.

Third, future studies should systematically incorporate domain-relevant moderators such as learner expertise, type and amount of variability ("noise"), dominant sensory components of the skill (visual, kinesthetic, proprioceptive), and contextual factors (task specificity, feedback type). Emphasis should also be placed on dynamic systems perspectives that account for non-linear performance trajectories and individual boundary conditions (Schöllhorn, 2000).

Finally, researchers should adopt a more reflective and interdisciplinary lens when interpreting motor learning outcomes. The theoretical foundations of the CI framework—particularly those invoking working memory mechanisms or elaboration/reconstruction processes—warrant re-examination in light of evolving neuroscientific and systems-based models (Ammar et al., 2023, 2024, Zou et al., 2025). For instance, the interpretation of working memory in CI theory often relies on Baddeley and Hitch's (1974) model, which was primarily based on visual-spatial and sequential tasks. Notably, Baddeley (2018) himself acknowledged that this model had not been validated for proprioceptive, kinesthetic, or tactile domains. Exploratory studies that combine behavioral metrics with neurophysiological tools (e.g., EEG, HRV) may offer richer insights into the underlying mechanisms of motor learning beyond behavioral outcomes alone (Ammar et al., 2023, 2024, Zou et al., 2025).

In summary, advancing CI and motor learning research requires moving beyond the proliferation of more fragmented meta-analyses but in building integrative, differentially informed, and practically relevant research programs that bridge evidence synthesis with theory development and real-world application.

Conclusion

The ongoing debate surrounding CI phenomenon continues to expose fundamental methodological and interpretive shortcomings in motor learning research. Although meta-analyses are often regarded as authoritative tools for synthesizing evidence, their reliability is contingent upon the robustness and transparency of their methodological foundations—an assumption that is frequently unfulfilled. In this critical examination, we have shown that the apparent divergence in findings between recent meta-analyses—particularly those by Ammar et al., (2023, 2024) and Czyż et al., (2024a, b)—is less a matter of conflicting data than of variability in conceptual frameworks, selective inclusion criteria, outcome definitions, and questionable analytic strategies, all of which may benefit from closer evaluation and further refinement.

Crucially, we have identified methodological oversights that could affect the validity of certain conclusions, including inadequate control for baseline equivalence, the exclusion of ecologically valid studies conducted in applied settings, and the unwarranted generalization of narrow findings into broader theoretical claims. Such practices are not merely technical flaws; they raise serious epistemological concerns. When a series of arguments is constructed to elevate the perceived integrity of a given synthesis—especially when these are based on misrepresentations or straw man distortions—one might reasonably ask whether rhetorical positioning is displacing scientific rigor.

Our analysis further warns of a growing trend in motor learning literature: the unchecked proliferation of meta-analyses that, rather than clarifying complex phenomena, may contribute to fragmentation, redundancy, and interpretive confusion. In a field already constrained by a scarcity of high-quality primary studies and considerable heterogeneity, such proliferation exacerbates the risk of misleading conclusions. We encourage researchers and reviewers to consider ways to minimise methodological shortcuts—such as ignoring baseline discrepancies, glossing over bias assessments, or drawing unwarranted generalizations—that could threaten both the scientific integrity and practical applicability of findings.

Moving forward, future motor-learning research could benefit from stronger methodological foundations—including robust randomized controlled trials, multilevel modeling strategies, and field-specific appraisal tools—and to shift from repetitive meta-analyses toward more integrative syntheses such as umbrella reviews. Ultimately, we advocate for a more context-aware, transparent, and theoretically grounded approach to evidence synthesis in motor learning—guided not by routine use of statistical tools, but by thoughtful methodological choices, clear theoretical frameworks, and careful attention to context, with the aim of producing meaningful insights that are appropriately and cautiously generalizable.

Author Contribution Conceptualization: AA, KT, AS, HJ, WIS; writing original draft: AA; writing—review and editing: AA, KT, AS, HAJ, WIS; validation: AA, KT, AS, HAJ, WIS; supervision: HJ, WIS.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors declare that this research received no specific grant from any funding agency and that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Data will be made available upon reasonable request.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Acree, M. C. (2021). The myth of statistical inference. *Springer Nature Switzerland*. https://doi.org/10.1007/978-3-030-73257-8

Afonso, J., Ramirez-Campillo, R., Clemente, F. M., Büttner, F. C., & Andrade, R. (2024). The perils of misinterpreting and misusing"publication bias" in meta-analyses: An education review on funnel plot-based methods. *Sports Medicine, 54*(2), 257–269. https://doi.org/10.1007/s40279-023-01927-9

Aguinis, H., Gottfredson, R. K., & Wright, T. A. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior,32*(8), 1033–1043. https://doi.org/10.1002/job.719

Ammar, A., & Schöllhorn, W. I. (2024). *Comment on Czyż et al. (2024) on contextual interference in motor learning* [Preprint]. SportRxiv. https://doi.org/10.51224/SRXIV.435

Ammar, A., Trabelsi, K., Boujelbane, M. A., Boukhris, O., Glenn, J. M., Chtourou, H., & Schöllhorn, W. I. (2023). The myth of contextual interference learning benefit in sports practice: A systematic review and meta-analysis. *Educational Research Review,39*, Article 100537. https://doi.org/10.1016/j.edurev.2023.100537

Ammar, A., Trabelsi, K., Boujelbane, M. A., Salem, A., Boukhris, O., Glenn, J. M., Zmijewski, P., Jahrami, H. A., Chtourou, H., & Schöllhorn, W. I. (2024). The effects of contextual interference learning on the acquisition and relatively permanent gains in skilled performance: A critical systematic review with multilevel meta-analysis. *Educational Psychology Review*.https://doi.org/10.1007/s10648-024-09892-z

Apidogo, J. B., Burdack, J., & Schöllhorn, W. I. (2021). Repetition without repetition or differential learning of multiple techniques in volleyball? *International Journal of Environmental Research and Public Health, 18*(19), 10499. https://doi.org/10.3390/ijerph181910499

Apidogo, J. B., Burdack, J., & Schöllhorn, W. I. (2022). Learning multiple movements in parallel—accurately and in random order, or each with added noise? *International Journal of Environmental Research and Public Health,19*(17), 10960. https://doi.org/10.3390/ijerph191710960

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology,12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Baddeley, A. D. (2018). *Exploring working memory: Selected works of Alan Baddeley*. Routledge. https://doi.org/10.4324/9781315111261-26

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1
Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of skill* (pp. 215–244). Academic Press.
Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In *Topics in learning and performance* (pp. 131–159). Academic Press.
Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 23–44). Lawrence Erlbaum.
Beik, M., & Fazeli, D. (2021). The effect of learner-adapted practice schedule and task similarity on motivation and motor learning in older adults. *Psychology of Sport and Exercise, 54*, 101911. https://doi.org/10.1016/j.psychsport.2021.101911
Beik, M., Taheri, H., Saberi Kakhki, A., & Ghoshuni, M. (2021). Algorithm-based practice schedule and task similarity enhance motor learning in older adults. *Journal of Motor Behavior, 53*(4), 458–470. https://doi.org/10.1080/00222895.2020.1797620
Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P., Lau, C. L. L., & Joo, H. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership–performance relationship. *Strategic Management Journal,37*(3), 477–497. https://doi.org/10.1002/smj.2338
Boyce, B. A., & Del Rey, P. (1990). Designing applied research in a naturalistic setting using a contextual interference paradigm. *Journal of Human Movement Studies, 18*, 189–200.
Brady, F. (2004). Contextual interference: A meta-analytic study. *Perceptual and Motor Skills, 99*(1), 116–126. https://doi.org/10.2466/pms.99.1.116-126
Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. *Quest, 50*(3), 266–293. https://doi.org/10.1080/00336297.1998.10484285
Broadbent, D. P., Causer, J., Williams, A. M., & Ford, P. R. (2017). The role of error processing in the contextual interference effect during the training of perceptual-cognitive skills. *Journal of Experimental Psychology: Human Perception and Performance, 43*(7), 1329–1342. https://doi.org/10.1037/xhp0000375
Buszard, T., Reid, M., Krause, L., Kovalchik, S., & Farrow, D. (2017). Quantifying contextual interference and its effect on skill transfer in skilled youth tennis players. *Frontiers in Psychology, 8*, 1931. https://doi.org/10.3389/fpsyg.2017.01931
Chan, M. E., & Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science, 7*(1), 79–92. https://doi.org/10.1177/1745691611429355
Czyż, S. H. (2025). Commentary on Ammar et al. (2024) "The effects of contextual interference learning on the acquisition and relatively permanent gains in skilled performance: Critical systematic review with multilevel meta-analysis." *Educational Psychology Review,37*, Article 28. https://doi.org/10.1007/s10648-025-10006-6
Czyż, S., & Wójcik, A. (2024). Rebuttal to Ammar and Schöllhorn on "Comment on Czyż et al. (2024) on contextual interference in motor learning" [Preprint]. *SportRxiv*. https://doi.org/10.51224/SRXIV.437
Czyż, S. H., Wójcik, A. M., Solarská, P., & Kiper, P. (2024a). High contextual interference improves retention in motor learning: Systematic review and meta-analysis. *Scientific Reports, 14*, 15974. https://doi.org/10.1038/s41598-024-65753-3
Czyż, S. H., Wójcik, A. M., & Solarská, P. (2024b). The effect of contextual interference on transfer in motor learning: A systematic review and meta-analysis. *Frontiers in Psychology,15*, Article 1377122. https://doi.org/10.3389/fpsyg.2024.1377122
Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2023). Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 6.4, updated August 2023). Cochrane. https://training.cochrane.org/handbook
Esposito, G., & Vivanti, G. (2013). Gross motor skills. In F. R. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders* (pp. –). Springer. https://doi.org/10.1007/978-1-4419-1698-3_179
Ewald, H., Klerings, I., Wagner, G., Heise, T. L., Stratil, J. M., Lhachimi, S. K., Hemkens, L. G., Gartlehner, G., Armijo-Olivo, S., & Nussbaumer-Streit, B. (2022). Searching two or more databases decreased the risk of missing relevant studies: A meta-research study. *Journal of Clinical Epidemiology,149*, 154–164. https://doi.org/10.1016/j.jclinepi.2022.05.022
Feest, U. (2019). Why replication is overrated. *Philosophy of Science, 86*(5), 895–905. https://doi.org/10.1086/705451

Fialho, J. V. A. P., Benda, R. N., & Ugrinotitsch, H. (2006). The contextual interference effect in a serve skill acquisition with experienced volleyball players. *Journal of Human Movement Studies, 50*, 65–77.

French, K. E., Rink, J. E., & Werner, P. H. (1990). Effects of contextual interference on retention of three volleyball skills. *Perceptual and Motor Skills, 71*(1), 179–186. https://doi.org/10.2466/pms.71.1.179-186

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science, 1*(2), 198–218. https://doi.org/10.1177/2515245918771329

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Glass, G. V. (2015). Meta-analysis at middle age: A personal history. *Research Synthesis Methods, 6*(3), 221–231. https://doi.org/10.1002/jrsm.1146

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*(7695), 175–182. https://doi.org/10.1038/nature25753

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills, 78*(3), 835–841. https://doi.org/10.2466/pms.78.3.835-841

Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: A practical guide. *Management Review Quarterly, 72*(1), 1–19. https://doi.org/10.1007/s11301-021-00247-4

Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly,94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Jeon, M. J., Jeon, H. S., Yi, C. H., Kwon, O. Y., You, S. H., & Park, J. H. (2021). Block and random practice: A Wii fit dynamic balance training in older adults. *Research Quarterly for Exercise and Sport,92*(3), 352–360. https://doi.org/10.1080/02701367.2020.1733456

Jiménez-Díaz, J., Morera-Castro, M., & Salazar, W. (2018). The contextual interference effect on the performance of fundamental motor skills in adults. *Human Movement,19*(1), 20–25. https://doi.org/10.5114/hm.2018.73608

Jones, L. L., & French, K. E. (2007). Effects of contextual interference on acquisition and retention of three volleyball skills. *Perceptual and Motor Skills, 105*(3), 883–890. https://doi.org/10.2466/pms.105.3.883-890

Klapp, S. T. (1995). Motor response programming during simple choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 1015–1027. https://doi.org/10.1037/0096-1523.21.5.1015

Klapp, S. T. (2003). Reaction time analysis of two types of motor preparation for speech articulation: Action as a sequence of chunks. *Journal of Motor Behavior, 35*(2), 135–150. https://doi.org/10.1080/00222890309602129

Krause, D., Brüne, A., Fritz, S., Kramer, P., Meisterjahn, P., Schneider, M., & Sperber, A. (2014). Learning of a golf putting task with varying contextual interference levels induced by feedback schedule in novices and experts. *Perceptual and Motor Skills, 118*(2), 384–399. https://doi.org/10.2466/23.30.PMS.118k17w3

Lage, G. M., Faria, L. O., Ambrósio, N. F. A., Borges, A. M. P., & Apolinário-Souza, T. (2021). What is the level of contextual interference in serial practice? A meta-analytic review. *Journal of Motor Learning and Development, 10*(1), 224–242. https://doi.org/10.1123/jmld.2021-0020

Lee, T. D., & Carnahan, H. (1990). Bandwidth knowledge of results and motor learning: More than just a relative frequency effect. *The Quarterly Journal of Experimental Psychology,42*(4), 777–789. https://doi.org/10.1080/14640749008401249

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg, & I. M. Franks (Eds.), *Differing perspectives in motor learning, memory, and control* (Advances in Psychology, Vol. 27, pp. 3–22). North-Holland. https://doi.org/10.1016/S0166-4115(08)62528-5

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,9*(4), 730–746. https://doi.org/10.1037/0278-7393.9.4.730

Li, Y., & Lima, R. P. (2002). Rehearsal of task variations and contextual interference effect in a field setting. *Perceptual and Motor Skills, 94*(3), 750–752. https://doi.org/10.2466/pms.2002.94.3.750

Lipton, P. (2000). Inference to the best explanation. In W. H. Newton-Smith (Ed.), *A companion to the philosophy of science* (pp. 184–193). Blackwell.

McKay, B., Hussien, J., Vinh, M.-A., Mir-Orefice, A., Brooks, H., & Ste-Marie, D. M. (2022a). Meta-analysis of the reduced relative feedback frequency effect on motor learning and performance. *Psychology of Sport and Exercise, 61*, 102165. https://doi.org/10.1016/j.psychsport.2022.102165

McKay, B., Yantha, Z., Hussien, J., Carter, M., & Ste-Marie, D. (2022b). Meta-analytic findings of the self-controlled motor learning literature: Underpowered, biased, and lacking evidential value. *Meta-Psychology, 6*. https://doi.org/10.15626/MP.2021.2803

McKay, B., Bacelar, M. F., Parma, J. O., Miller, M. W., & Carter, M. J. (2023). 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. *International Review of Sport and Exercise Psychology*. https://doi.org/10.1080/1750984X.2023.2207255

Memmert, D. (2006). Long-term effects of type of practice on the learning and transfer of a complex motor skill. *Perceptual and Motor Skills, 103*(3), 912–916. https://doi.org/10.2466/pms.103.3.912-916

Menayo, R., Sabido, R., Fuentes, J. P., Moreno, F. J., & Garcia, J. A. (2010). Simultaneous treatment effects in learning four tennis shots in contextual interference conditions. *Perceptual and Motor Skills,110*(2), 661–673. https://doi.org/10.2466/pms.110.2.661-673

Moretto, N. A., Marcori, A. J., & Okazaki, V. H. A. (2018). Contextual interference effects on motor skill acquisition, retention and transfer in sport rifle shooting. *Human Movement, 19*(1), 99–104.

National Institutes of Health. (2001, October). *NIH policy and guidelines on the inclusion of women* and *minorities as subjects in clinical research* (amended October 2001). https://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm

North, J. S., Bezodis, N. E., Murphy, C. P., Runswick, O. R., Pocock, C., & Roca, A. (2019). The effect of consistent and varied follow-through practice schedules on learning a table tennis backhand. *Journal of Sports Sciences, 37*(6), 613–620. https://doi.org/10.1080/02640414.2018.1522683

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 1–11. https://doi.org/10.1186/s13643-021-01626-4

Pham, M. T., & Oh, T. T. (2021). Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology, 31*(1), 163–176. https://doi.org/10.1002/jcpy.1209

Pollatou, E., Kioumourtzoglou, E., Agelousis, N., & Mavromatis, G. (1997). Contextual interference effects in learning novel motor skills. *Perceptual and Motor Skills, 84*(2), 487–496. https://doi.org/10.2466/pms.1997.84.2.487

Rendell, M. A., Masters, R. S. W., Farrow, D., & Morris, T. (2010). An implicit basis for the retention benefits of random practice. *Journal of Motor Behavior, 43*(1), 1–13. https://doi.org/10.1080/00222895.2010.530304

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., et al. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews, 10*, 39. https://doi.org/10.1186/s13643-020-01542-z

Rothman, K. J., Greenland, S., & Lash, T. L. (2008).*Modern epidemiology* (3rd ed.). Lippincott Williams & Wilkins.

Sadri, K., Mohommadzadeh, H., & Khani, M. (2013). The effect of contextual interference on acquisition and learning badminton skills among children aged 10 to 12. *Annals of Applied Sport Science, 1*(3), 39–46.

Sattelmayer, M., Elsig, S., Hilfiker, R., & Baer, G. (2016). A systematic review and meta-analysis of selected motor learning principles in physiotherapy and medical education. *BMC Medical Education,16*, 15. https://doi.org/10.1186/s12909-016-0538-z

Schmidt, R. A., & Lee, T. D. (2019). *Motor learning and performance* (6th ed.). Human Kinetics

Schmidt, R. A., Lee, T. D., Winstein, C. J., Wulf, G., & Zelaznik, H. N. (2011). *Motor control and learning: A behavioral emphasis* (5th ed.). Human Kinetics

Schöllhorn, W. I. (2000). Application of systems dynamic approach to technique and strength training. *Acta Academiae Olympiquae estoniae, 8*, 67–85.

Schöllhorn, W. I. (2016). Invited commentary: Differential learning is different from contextual interference learning. *Human Movement Science, 47*, 240–245. https://doi.org/10.1016/j.humov.2015.11.018

Schöllhorn, W. I., Rizzi, N., Slapšinskaitė-Dackevičienė, A., & Leite, N. (2022). Always pay attention to which model of motor learning you are using. *International Journal of Environmental Research and Public Health, 19*(2), 711. https://doi.org/10.3390/ijerph19020711

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical research ed.),340*, c332. https://doi.org/10.1136/bmj.c332

Shan, Y., & Williamson, J. (2022). Evidential monism, evidential pluralism, or evidential contextualism? An introduction to evidential diversity in the social sciences. *Synthese, 200*, 321. https://doi.org/10.1007/s11229-022-03801-z

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*(2), 179–187. https://doi.org/10.1037/0278-7393.5.2.179

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical research ed.),358*, Article j4008. https://doi.org/10.1136/bmj.j4008

Smith, P. J. K., & Davies, M. (1995). Applying contextual interference to the Pawlata roll. *Journal of Sports Sciences, 13*(6), 455–462. https://doi.org/10.1080/02640419508732262

Smith, P. J. K., & Rudisill, M. E. (1993). The influence of proficiency level, transfer distality, and gender on the contextual interference effect. *Research Quarterly for Exercise and Sport, 64*(2), 151–157. https://doi.org/10.1080/02701367.1993.10608792

Steele, J., Fisher, J. P., Smith, D., Schoenfeld, B. J., Yang, Y., & Nakagawa, S. (2023). Meta-analysis of variation in sport and exercise science: Examples of application within resistance training research. *Journal of Sports Sciences, 41*(17), 1617–1634. https://doi.org/10.1080/02640414.2023.2286748

Tsutsui, S., Satoh, M., & Yamamoto, K. (2013). Contextual interference modulated by pitcher skill level. *International Journal of Sport and Health Science*. https://doi.org/10.5432/ijshs.201211

van Wonderen, E., Zondervan-Zwijnenburg, M., & Klugkist, I. (2024). Bayesian evidence synthesis as a flexible alternative to meta-analysis: A simulation study and empirical demonstration. *Behavior Research Methods, 56*(4), 4085–4102. https://doi.org/10.3758/s13428-024-02350-2

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. https://doi.org/10.1007/BF02294384

van den Akker, O. R., van Assen, M. A. L. M., Bakker, M., Elsherif, M., Wong, T. K., & Wicherts, J. M. (2024). Preregistration in practice: A comparison of preregistered and non-preregistered studies in psychology. *Behavior Research Methods,56*(6), 5424–5433. https://doi.org/10.3758/s13428-023-02277-0

Wimsatt, W. C. (2012). Robustness, reliability, and overdetermination (1981). In L. Soler, E. Trizio, T. Nickles, spsampsps W. Wimsatt (Eds.), *Characterizing the robustness of science* (Boston Studies in the Philosophy of Science, Vol. 292, pp. 23–41). Springer. https://doi.org/10.1007/978-94-007-2759-5_2

Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination. In M. B. Brewer & B. E. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124–163). Jossey-Bass.

Wulf, G., & Lee, T. D. (1993). Contextual interference in movements of the same class: Differential effects on program and parameter learning. *Journal of Motor Behavior, 25*(4), 254–263. https://doi.org/10.1080/00222895.1993.9941646

Zou, L., Zhang, Z., Mavilidi, M., Chen, Y., Herold, F., Ouwehand, K., & Paas, F. (2025). The synergy of embodied cognition and cognitive load theory for optimized learning. *Nature Human Behaviour, 9*, 877–885. https://doi.org/10.1038/s41562-025-02152-2

Authors and Affiliations

Achraf Ammar[1,2,3,5] · Khaled Trabelsi[3,4,5] · Atef Salem[1,3] · Haitham Jahrami[6,7] · Wolfgang I. Schöllhorn[1]

✉ Achraf Ammar
acammar@uni-mainz.de; ammar.achraf@ymail.com

1 Department of Training and Movement Science, Institute of Sport Science, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany

2 Research Laboratory, Molecular Bases of Human Pathology, LR19ES13, Faculty of Medicine of Sfax, University of Sfax, Sfax 3029, Tunisia

3 High Institute of Sport and Physical Education, University of Sfax, Sfax, Tunisia

4 Research laboratory, Education, Motricity, Sport and Health (EM2S), LR15JS01 High Institute of Sport and Physical Education, University of Sfax, Sfax, Tunisia

5 Department of Movement Sciences and Sports Training, School of Sport Science, the University of Jordan, Amman, Jordan

6 Government Hospitals, P.O. Box 12, Manama, Bahrain

7 Department of Psychiatry, College of Medicine and Health Sciences, Arabian Gulf University, P.O. Box 26671, Manama, Bahrain