Data irregularities across five implicit learning articles by the same researchers

Brad McKay[1] & Michael J. Carter[1]

[1] Department of Kinesiology
McMaster University

Author Note

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

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

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

Abstract

We present a critical re-analysis of five implicit learning papers published by the same authors between 2010 and 2021. We calculated effect sizes for each pairwise comparison reported in the papers using the data published in each article. We further identified mathematically impossible data reported in multiple papers, either with deductive logic or by conducting a GRIMMER analysis of reported means and standard deviations. We found the pairwise effect sizes were implausible in all five articles in question, with Cohen's d values often exceeding 100 and sometimes exceeding 1000. Impossible statistics were reported in 4 out of the 5 articles. Reported test statistics and η^2 values were also implausible, with several $\eta^2 = .99$ and even $\eta^2 = 1.0$ for between-subjects main effects. The results reported in the five articles in question are unreliable. Many of the problems we identified could be spotted without further analysis, highlighting the need for adequate statistical training in the field of motor learning.

Keywords: Meta-science; GRIMMER; Effect sizes; Perceptual motor learning

Word count: X

Data irregularities across five implicit learning articles by the same researchers

Statistical reporting errors may commonly occur in psychology articles (Brown & Heathers, 2017; Nuijten et al., 2016) and such errors are often consistent with hypothesized results (Bakker & Wicherts, 2011). When the primary conclusions in research articles depend on reporting errors, replicability is unlikely and future research may be wasted if researchers attempt to build on the erroneously reported results (Munafò et al., 2017). In this paper, we scrutinize five papers published by the same two authors[1], where the authors report a high number of erroneous or implausible data on which their primary conclusions depend. We first became aware of the Lola and Tzetzis (2021) paper when the paper was highlighted in a social media post (Gray, 2021). During an initial read through by one of us (BM), a number of reporting and/or statistical issues were noticed. The paper also referenced past research published by these authors, and given our cause for concern over the issues found in the Lola and Tzetzis (2021) paper, we deemed it necessary to examine these other papers. We remain agnostic to the sources of error in each of these papers. Nevertheless, the data irregularities we found are similar across the target articles and at times even include repeated values (e.g., F-statistics) across multiple papers. Regardless of the conclusion one reaches with respect to the mechanism behind these errors, it is our contention that the results reported in these papers are unreliable and that the respective journals in which the papers are published should take corrective actions. Below, we outline our causes for concern and the overarching issues we found across the five articles in question.

The articles in question

We reanalyzed five articles by Afroditi Lola, George Tzetzis, and their colleague (Lola et al., 2012; Lola & Tzetzis, 2021, 2020; Tzetzis & Lola, 2010, 2015). All five papers described experiments that sampled young females who were enrolled in a volleyball camp (see Table 1). In their experiments, the authors investigated the effects of implicit and explicit

[1] One of the five papers had a third author.

instructions on learning motor and perceptual skills. Our reanalysis of the target articles evaluated the plausibility of the reported means, standard deviations, and test statistics. The five articles will be referred to throughout this paper using the following numbering system:

1. Lola, A.C., & Tzetzis, G.C. (2021). The effect of explicit, implicit and analogy instruction on decision making skill for novices, under stress. *International Journal of Sport and Exercise Psychology*, 1-21. https://doi.org/10.1080/1612197X.2021.1877325

2. Lola, A.C., & Tzetzis, G.C. (2020). Analogy versus explicit and implicit learning of a volleyball skill for novices: The effect on motor performance and self-efficacy. *Journal of Physical Education and Sport*, 20(5), 2478-2486. https://doi.org/10.7752/jpes.2020.05339

3. Tzetzis, G.C., & Lola, A.C. (2015). The effect of analogy, implicit, and explicit learning on anticipation in volleyball serving. *International Journal of Sport Psychology*, 46(2), 152-166. https://doi.org/10.7352/IJSP.2015.46.152

4. Lola, A.C., & Tzetzis, G.C., & Zetou, H. (2012). The effect of implicit and explicit practice in the development of decision making in volleyball serving. *Perceptual and Motor Skills*, 114(2), 665-678. https://doi.org/10.2466/05.23.25.PMS.114.2.665-678

5. Tzetzis, G.C., & Lola, C.A. (2010). The role of implicit, explicit instruction and their combination in learning anticipation skill, under normal and stress conditions. *International Journal of Sport Sciences and Physical Education*, 1, 54-59.[2]

[2] This article is in a journal of a publishing group that has been identified as a potential predatory journal. Further, we were unable to find an archived version of this article on the journal's webpage and were only able to find a posted version on ResearchGate (https://www.researchgate.net/publication/341001393_THE_ROLE_OF_IMPLICIT_EXPLICIT_INSTRUCTION_AND_THEIR_COMBINATION_IN_LEARNING_ANTICIPATION_SKILL_UNDER_NORMAL_AND_STRESS_CONDITIONS). In fact, the earliest available issue on the journal's webpage in from 2016.

Table 1

Participant demographics in each of the target articles.

Target article	Sample size and participant details
Article 1	
Lola & Tzetzis (2021)	60 females, age range: 10 to 11 years (M_{age} = 10.48, SD = 0.911)[a]
Article 2	
Lola & Tzetzis (2020)	80 females, age range: 10 to 11 years (M_{age} = 10.48, SD = 0.911)[a]
Article 3	
Tzetzis & Lola (2015)	60 females, age range: 9 to 12 years (M_{age} = 10.48, SD = 0.91)[a]
Article 4	
Lola et al (2012)	60 females, age range: 10 to 12 years (M_{age} = 11.2, SD = 0.3)
Article 5	
Tzetzis & Lola (2010)	48 females, age range: 12 to 13 years (M_{age} = 12.38, SD = 0.34)

Note. [a] Articles 1-3 report identical means and standard deviations for the age of their participants despite a different sample size in Article 2 from Articles 1 and 3, and a different age range in Article 3 from Articles 1 and 2.

Although there were some differences between the reported experiments in the target articles, there were many methodological commonalities that can be summarized. All five articles involved female children learning a volleyball skill as part of a volleyball camp. In each case, the participants were reported to have minimal experience (i.e., were described as novices) with the task at hand. The purpose of all five experiments was to evaluate perceptual or motor learning differences as a function of the type of instruction received during practice. Each experiment included a pre-test, an acquisition (i.e., practice) phase involving 12 sessions spaced over four weeks, and a post-test. The authors also included a high stress test in Articles 1 and 5.

In Articles 1-3, the groups differed with respect to the type of instruction received:

implicit, explicit, or analogy. In Articles 4 and 5, a sequential group (see below for description) replaced the analogy group. All five experiments also included a control group that did not practice the task. Implicit instruction did not contain any explicit information for how to perform the task and the learners were asked to perform a distracting task like counting backwards while practicing to prevent them from acquiring declarative rules for performance. In contrast, explicit instruction consisted of direct verbal instructions for performing the task. Analogy instruction was considered a type of implicit instruction wherein an analogy or metaphor was provided to the learner. For example, “Imagine that the opponents’ surface is covered with water. Send the ball where there is more water and no opponents at the court.” (Lola & Tzetzis, 2021, p. 9). Sequential instruction involved receiving explicit instruction for the first half of training followed by implicit instruction for the second half of training. Across experiments, the authors predicted that implicit forms of instruction—implicit, analogy, and sequential—would be more effective than explicit instruction for motor and perceptual learning. This advantage was also predicted to be greater when testing was conducted in a high stress situation. In Lola and Tzetzis (2021) for instance, high stress was induced by falsely telling participants that the best performers would be selected for a draft to the national team. Further, it was often predicted that analogy or sequential instruction would offer improvements relative to implicit instruction.

The primary outcome measures used in these experiments were reaction time (Articles 1, 3, 4, and 5), response accuracy (Articles 1, 3, 4), and motor performance measured on a 4-point scale (Article 2). In addition, Articles 1 and 5 included a measure of state anxiety with the Competitive State Anxiety Inventory-2 (CSAI-2) (Tsorbatzoudis et al., 1998), and Article 2 had a measure of self-efficacy using a Likert scale. Articles 1, 3, 4, and 5 also analyzed the number of explicit rules recalled.

Methods

None of the five articles in question included a link to a public repository where the data could be accessed. We first wrote (email sent February 11 2021) the corresponding author for Article 1 and asked if they would be willing to share the data for this experiment. The authors' response was that the data could not be shared as they were not finished with their analysis and were in the process of running different tests (A. Lola, personal communication, February 12 2021). We followed up this email (sent February 12 2021) by asking whether they would instead be willing to share the data from Articles 2 through 5 as these articles were less recent and presumably, all planned analyses had been completed. After a 2 week period with no response, we followed up on with a third email (sent February 26 2021) and reiterated our interest in obtaining their data from Articles 2 through 5. The authors' response was that they were unable to share data from any of these articles because in some cases they no longer had the data and in other cases they had plans to conduct further analyses (A. Lola, personal communication, March 2 2021).

Our first two requests did not include any indication that we were concerned about irregularities. Subsequently, in a fourth email (sent April 12 2021) we outlined our concerns for each article and once again reiterated our request to the authors to share any available data for any of the target articles. These requests were once again refused. The authors did address some specific concerns regarding Article 1 and provided some more general responses to our concerns; however, elements of their responses were inaccurate relative to the target articles. The authors did admit that some of the values reported in the target articles were incorrect but did not identify which values or articles. Despite this, the authors maintained that the data irregularities—identified in our email and described in this paper—do not impact the veracity of their analyses or conclusions (A. Lola, personal communication, April 22 2021). We illustrate below that the data *and* analyses reported in each of the articles reviewed are unreliable. Our extracted data and analysis scripts can be accessed at the following link: [**LINK**].

Effect size calculations and simulations

Means and standard deviations were extracted from each article for all measures and time points that were reported. Cohen's d was calculated for each pairwise comparison using the `compute.es` package in `R`. To provide context, we simulated data with true effect sizes of $d = .80$ and $d = 3.0$ one million times each and report the range of effect sizes observed in those simulations.

Mathematically impossible data and granularity analysis

In two of the articles in question, it was clear that the some of the reported results were not mathematically possible based on the scale of measurement that was used. When outcomes were single item integers, such as the number of explicit rules recalled, we used a web application (http://www.prepubmed.org/grimmer_sd/) to conduct a granularity analysis called GRIMMER (Anaya, 2016). Using GRIMMER it is possible to determine if specific means and standard deviations are possible for a given sample size. To be conservative, we specified that we did not know whether the standard deviation was calculated for the sample or population, nor whether ambiguous values were rounded up or down. Mean and standard deviation pairs that are mathematically possible are considered GRIMMER consistent, while mean and standard deviation pairs that are not mathematically possible are GRIMMER inconsistent.

Test statistics and η^2

Each of the articles reported only omnibus test statistics and then reported post-hoc analyses with symbols demarcating significant and non-significant differences. In response to our expressions of concern, the authors suggested that many of the issues were due to misprints in the articles. Specifically, they indicated that the reported means and standard deviations in their tables were incorrect and the root of the errors had to be from them outsourcing the formatting of their tables. The authors then insisted that despite these typographic errors, their discussion of the results and corresponding conclusions were still accurate (A. Lola, personal communication, April 22, 2021). However, the test statistics reported for many analyses were implausibly large and the authors often reported η^2 values associated with the omnibus test. Our examination of the reported η^2 values revealed that many were not only implausibly large, but also consistent with the reported pairwise effects.

Results

Implausible effect sizes

Cohen's d is used to describe the standardized mean difference of an effect and values can range between 0 and infinity. Cohen's d_s (Cohen, 1988) is the observed difference between group means divided by their pooled standard deviation (see Lakens, 2013 for a detailed discussion). Conventional benchmarks for small, medium and large effects are $d =$ .2, .5, and .8, respectively (Cohen, 1962); however, this *cargo-cult* approach to effect size interpretation has been heavily discouraged (Correll et al., 2020; Field & Iles, 2016; Lakens, 2013; Thompson, 2007). Recently, an analysis of 6447 Cohen's d statistics extracted from social psychology meta-analyses observed median and 75th percentile Cohen's d values of .36 and .65, respectively—suggesting the conventional benchmarks may overestimate typical effects (Lovakov & Agadullina, 2021).

To evaluate the maximum range of plausible Cohen's d statistics one might encounter from experiments similar to those reported in Articles 1-5, we conducted two simulations that each consisted of 1,000,000 experiments. For the first simulation we set the true effect size at $d = .80$, the conventional benchmark for a "large" treatment effect. The largest effect size observed from the 1,000,000 simulated experiments was $d = 2.97$ (see Figure 1). In the second simulation, we set the true effect size at $d = 3.0$, an unrealistically large effect size that might rarely be encountered in the psychology literature. The maximum effect size observed in the 1,000,000 simulated experiments was $d = 6.6$.

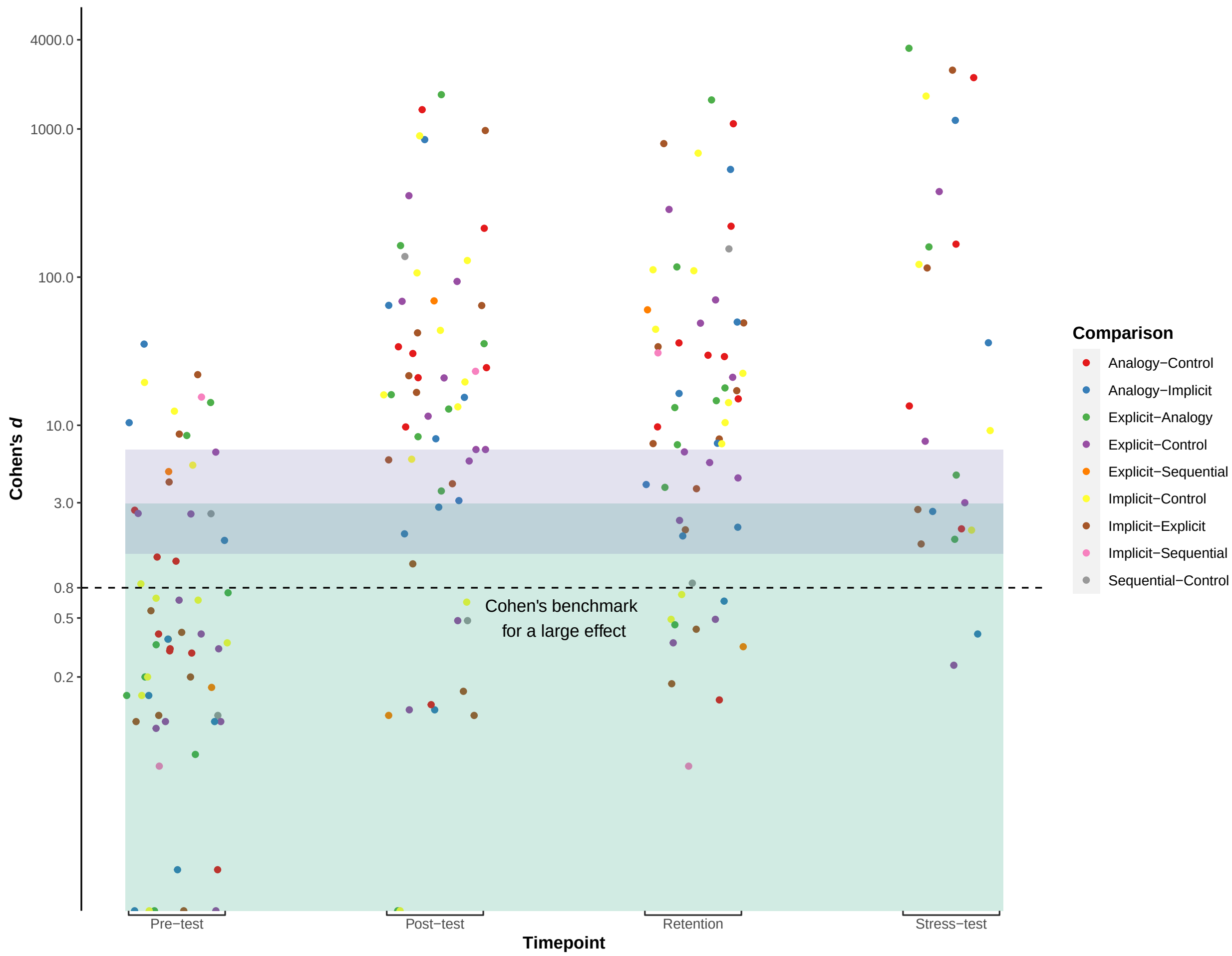

Figure 1

Absolute Cohen's d estimates from Articles 1-4 plotted on a logarithmic scale. All pairwise comparisons have been included for all dependent measures in each experiment. The range of observed values from a simulation of 1,000,000 experiments with a true effect of d = .80 is illustrated by light blue and dark blue regions of the figure, reaching a maximum value of d = 2.97. The range of observed values from a simulation of 1,000,000 experiments with a true effect of d = 3.0 is illustrated by the purple and dark blue regions of the figure, reaching a maximum value of d = 6.86.

In the context of the maximum values observed in our simulations, all five articles in question reported implausibly large effect sizes. In Article 1, the smallest pre-test difference on the reaction measure was $d = 1.29$, while the largest pre-test difference was $d = 35.32$, although none of the groups were reported as significantly different in the article. The smallest post-intervention effect at any of the three time points was $d = 286.42$, while the largest effect was $d = 3504.86$.

A similar picture emerges when analyzing the accuracy data. All the pre-test differences were implausibly large (all d's ≥ 2.52) despite being reported as not significantly different in the articles. Ten of the pairwise comparisons resulted in d's ≥ 100 following treatment with the independent variables. The motor component data also revealed many post-treatment effect sizes that are implausibly large, ranging from $d = 1.16$ to $d = 13.5$.

In Article 2, post-intervention motor performance effect sizes ranged from $d = 3.1$ to $d = 20.95$. Similarly, post-intervention self-efficacy effect sizes ranged from $d = 1.79$ to $d = 44.46$. Similarly, in Article 3 post-intervention reaction time effect sizes ranged from $d = 2.28$ to $d = 35.97$. Continuing this pattern, post-intervention response accuracy effect sizes ranged from $d = 5.84$ to $d = 29.7$.

In Article 4, many response accuracy effect sizes were implausibly large beginning at pre-test, wherein effects ranged from $d = 2.53$ to $d = 15.50$. Nevertheless, all pre-test comparisons were reported as non-significant. Following intervention, the effect sizes ranged from $d = 23.13$ to $d = 155.08$.

The reaction time effect sizes were not implausibly large at any time point, ranging from $d = 0.0$ to $d = .86$. However, the authors reported an implausibly large effect size for the 4 (Group) x 3 (Time) ANOVA, $\eta^2 = .94$. Further, despite only one pairwise comparison being statistically significant, all post-intervention comparisons were reported as being significant in the article.

In Article 5 the authors did not report means and standard deviations for most of the analyses. However, η^2 effect sizes were reported and these ranged from $\eta^2 = .52$ to $\eta^2 = .98$. These effect sizes are discussed further below.

All the post-intervention effects reviewed above were directionally consistent with the researchers' expectations. The sometimes implausibly large pre-test effects were not expected, but also were not reported as significant.

Impossible data and granularity analysis

In Article 1, the CSAI-2 was used to assess the level of cognitive and somatic stress experienced by participants. Responses were measured on a Likert scale ranging from 1 to 4 with the data appearing to represent the average response per item. At each of the three low-stress time points, the means reported for all four groups ranged from 1.02 to 1.09. During the high-stress time point, the means ranged from 3.95 to 4.09. The means for two groups were reported as greater than 4, which is not possible given the maximum score on the CSAI-2 is 4.

In Article 2, participants were asked to receive a served volleyball and pass it to a target consisting of three concentric circles. Motor performance was measured based on where the pass landed, with three points awarded for a pass to the central circle on the target, two points for the middle circle, one point for the outermost circle, and zero points for a pass that missed the target.[3] Results were presented as average performance per trial and the analogy group was reported to have a mean score of 3.00 at retention (a perfect score) but with a standard deviation of .09. The perfect score was not a rounding error because the same group was reported to have a mean score of 2.99 with a standard deviation of .11 on the post-test. These data are not possible.

[3] Independent of the issues we have raised, this approach to measuring motor performance has been shown to be inappropriate and flawed for this type of task (Fischman, 2015; Hancock et al., 1995; Reeve et al., 1994).

Articles 1, 4, and 5 reported the means and standard deviations for the number of explicit rules recalled by participants following the intervention phase. As a single item analysis of integers these results were suitable for a GRIMMER analysis.

In Article 1, the mean and standard deviation pairs were GRIMMER inconsistent for all four groups. In Article 4, the mean and standard deviation pair was GRIMMER consistent for the explicit rules group (M = 4.8, SD = 1.78). The mean and standard deviation pairs for the remaining three groups were GRIMMER inconsistent. In Article 5, the mean and standard deviation pairs were GRIMMER consistent for three of the four groups if the standard deviations were calculated for the population rather than the sample. For two of the groups, however, they were consistent regardless which method of calculating the standard deviation was used. However, the results for the explicit group were GRIMMER inconsistent (M = 4.8, SD = 1.78).[4]

Test statistics and η^2

Eta-squared (η^2) is calculated by dividing the sum of squares for the effect by the total sum of squares. It can be interpreted as analogous to R^2: It represents the total variation in the dependent measure that can be explained by a given main effect or interaction in an analysis of variance (Lakens, 2013). Benchmarks have been suggested for small, medium, and large η^2 effect sizes as $\eta^2 = .01$, .06, and .14, respectively (Cohen, 1988). Importantly, if a main effect of instruction type results in $\eta^2 = .99$, this suggests that 99% of the total variability in outcome measure can be explained by group assignment alone. Such a result is implausible.

The range of significant F-statistics reported across the five articles in question was

[4] You may have noticed that the same mean and standard deviation pairing (M = 4.8, SD = 1.78) was classified as GRIMMER inconsistent for one paper and consistent for the other. This is because of sample size differences ($n = 20$ for the consistent result and $n = 12$ for the inconsistent one).

$F = 28.60$ to $F = 101{,}489.71$. Only three F-statistics were under 100 (28.60, 30.73, and 91.75), and all three were for interactions.

Article 1 did not report η^2 values but had the largest pairwise effects and F-statistics of the five articles in question. Article 2 reported $\eta^2 = .994$, $\eta^2 = .996$, and $\eta^2 = .996$ for the Time, Group, and Time x Group effects on motor performance. Similarly, variance explained on the self-efficacy measure was $\eta^2 = .995$, $\eta^2 = .994$, $\eta^2 = .997$ for the Time, Group, and Time x Group effects, respectively. Article 3 also reported $\eta^2 = .99$ for all three effects on both response time and response accuracy measures.

Article 4 reported $\eta^2 = 1.0$ for the main effect of Time and the Time x Group interaction, as well as $\eta^2 = .95$ for the main effect of Group on the response time measure. Interestingly, the Time x Group interaction had the smallest reported significant F-statistic among the five articles in question. With respect to response accuracy, the reported effects were $\eta^2 = .98$, $\eta^2 = .94$, $\eta^2 = .93$ for the Time, Group, and Time x Group analyses. Article 5 reported $\eta^2 = .66$, $\eta^2 = .52$, $\eta^2 = .72$ for the Time, Group, and Time x Group analyses.

Other oddities

Although the means and standard deviations for the explicit rules analysis were only reported in three of the articles in question, analyses were reported in Articles 1, 3, 4, and 5. The reported test statistic in these four articles was $F = 52.67$, albeit with different degrees of freedom in Article 5 that reflected the different sample size in this experiment (48 versus 60 in Articles 1, 3, and 4). Articles 1-3 were published over a span of 6 years with reported samples sizes of 60 in Articles 1 and 3, and 80 in Article 2. Yet, the authors report identical means and standard deviations for the age of their participants in these three articles (see Table 1). We assumed that each article was based on different samples as none of the articles mentioned using any previously published data.

Discussion

We have reviewed concerning data irregularities spanning five articles investigating implicit motor and perceptual learning (Lola et al., 2012; Lola & Tzetzis, 2021, 2020; Tzetzis & Lola, 2010, 2015). These data irregularities include implausibly large effect sizes for pairwise comparisons and impossible descriptive statistics—both of which have been acknowledged by the authors as misprints due to an outsourcing of table formatting (A. Lola, personal communication, April 22, 2021). Further, the reported test statistics and associated η^2 values are also implausibly large, which is inconsistent with the authors' claim that their results and discussions remain valid despite these aforementioned typographic errors in the tables. Considering these findings, it seems the conclusions from these articles are not reliable.

It is noteworthy that the results reported in each of these articles perfectly reflect the authors' expectations. Indeed, our attention was drawn to these articles after the Lola and Tzetzis (2021) paper was shared on Twitter (Gray, 2021), possibly because the results appeared to be exemplary. Although these errors seem unlikely to have aligned with expectations by chance alone, our exposure to them occurred after they had been selected for publication. We cannot rule out that these papers were selected for publication because of exemplary results and happened to have errors, and this selection caused those errors to correlate with the authors' expectations.

Other irregularities, such a repeating F-statistic for all four analyses of explicit rules and the recurring age of participants, potentially reflect sloppiness more than expectation. Indeed, the authors have already admitted that some values reported in their tables were in error, and it seems errors occurred in each of the articles we have reviewed. These errors were pervasive and appear to have substantially affected the conclusions of the articles in question. At a minimum, the consistent reporting errors across these five articles seem to reflect excessive carelessness throughout the publication process. Even if the authors offer corrections, which they have suggested they intend to do[5], many in the research community may find it difficult to trust any of these results.

Data, materials, and code availability

All material, data, and scripts to reproduce our analyses and figure can be accessed here: [**LINK**].

R packages used in this project

R [Version 4.1.0; R Core Team (2021)] and the R-packages *compute.es* [Version 0.2.5; Re (2013)], *kableExtra* [Version 1.3.4; Zhu (2021)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], and *tidyverse* [Version 1.3.1; Wickham et al. (2019)].

Conflict of interest

The authors declare no competing interests.

[5] As of today's date (2021-07-08), there is no indication that such corrective actions have been taken by the authors.

References

Anaya, J. (2016). *The GRIMMER test: A method for testing the validity of reported measures of variability*. https://doi.org/10.7287/peerj.preprints.2400v1

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. https://github.com/crsh/papaja

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*(3), 666–678. https://doi.org/10.3758/s13428-011-0089-5

Brown, N. J. L., & Heathers, J. A. J. (2017). The GRIM Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology. *Social Psychological and Personality Science*, *8*(4), 363–369. https://doi.org/10.1177/1948550616673876

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal and Social Psychology*, *65*, 145–153. https://doi.org/10.1037/h0045186

Correll, J., Mellinger, C., McClelland, G. H., & Judd, C. M. (2020). Avoid cohen's 'small', 'medium', and 'large' for power analysis. *Trends in Cognitive Sciences*, *24*(3), 200–207. https://doi.org/10.1016/j.tics.2019.12.009

Field, A. P., & Iles, J. (2016). *An adventure in statistics: The reality enigma*.

Fischman, M. G. (2015). On the continuing problem of inappropriate learning measures: Comment on Wulf et al. (2014) and Wulf et al. (2015). *Human Movement Science*, *42*, 225–231. https://doi.org/10.1016/j.humov.2015.05.011

Gray, R. [@ShakeyWaits]. (2021). *The effect of explicit, implicit and analogy instruction on decision making skill for novices, under stress.* [Tweet; Thumbnail link to article]. Twitter. https://twitter.com/ShakeyWaits/status/1359501377073012736.

Hancock, G. R., Butler, M. S., & Fischman, M. G. (1995). On the problem of two-dimensional error scores: Measures and analyses of accuracy, bias, and consistency. *Journal of Motor Behavior*, *27*(3), 241–250. https://doi.org/10.1080/00222895.1995.9941714

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lola, A. C., & Tzetzis, G. C. (2021). The effect of explicit, implicit and analogy instruction on decision making skill for novices, under stress. *International Journal of Sport and Exercise Psychology*, *0*(0), 1–21. https://doi.org/10.1080/1612197X.2021.1877325

Lola, A. C., & Tzetzis, G. C. (2020). Analogy versus explicit and implicit learning of a volleyball skill for novices: The effect on motor performance and self-efficacy. *Journal of Physical Education and Sport*, *20*(5), 2478–2486. https://www.cabdirect.org/cabdirect/abstract/20203562097

Lola, A. C., Tzetzis, G. C., & Zetou, H. (2012). The Effect of Implicit and Explicit Practice in the Development of Decision Making in Volleyball Serving. *Perceptual and Motor Skills*, *114*(2), 665–678. https://doi.org/10.2466/05.23.25.PMS.114.2.665-678

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

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

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (19852013). *Behavior Research Methods*, *48*(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

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

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

Reeve, T. G., Fischman, M. G., Christina, R. W., & Cauraugh, J. H. (1994). Using one-dimensional task error measures to assess performance on two-dimensional tasks: Comment on 'attentional control, distractors, and motor performance'. *Human Performance*, *7*(4), 315–319. https://doi.org/10.1207/s15327043hup0704_6

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, *44*(5), 423–432. https://doi.org/10.1002/pits.20234

Tsorbatzoudis, H., Barkoukis, V., Kaissidis-Rodafinos, A., & Grouios, G. (1998). A test of the reliability and factorial validity of the greek version of the CSAI-2. *Research Quarterly for Exercise and Sport*, *69*(4), 416–419. https://doi.org/10.1080/02701367.1998.10607717

Tzetzis, G. C., & Lola, A. C. (2010). The role of implicit, explicit instruction and their combination in learning anticipation skill under normal and stress conditions. *International Journal of Sport Sciences and Physical Education*, *1*, 54–59.

Tzetzis, G. C., & Lola, A. C. (2015). The effect of analogy, implicit, and explicit learning on anticipation in volleyball serving. *International Journal of Sport Psychology*, *46*(2), 152–166. https://doi.org/10.7352/IJSP.2015.46.152

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

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax.* https://CRAN.R-project.org/package=kableExtra