Citation for published version:
Brown, DJ & Fletcher, D 2017, 'Effects of psychological and psychosocial interventions on sport performance: a meta-analysis', Sports Medicine , vol. 47, no. 1, pp. 77-99. https://doi.org/10.1007/s40279-016-0552-7

DOI:
10.1007/s40279-016-0552-7

Publication date:
2017

Document Version
Peer reviewed version

Link to publication

Effects of Psychological and Psychosocial Interventions on Sport Performance: A Meta-Analysis

Daniel J. Brown and David Fletcher

Loughborough University, United Kingdom

The final publication is available at Springer via

http://dx.doi.org/10.1007/s40279-016-0552-7.

Author Note

Daniel J. Brown and David Fletcher, School of Sport, Exercise and Health Sciences, Loughborough University, United Kingdom.

Daniel J. Brown is now at the Department for Health, University of Bath, United Kingdom.

Correspondence concerning this article should be addressed to Daniel J. Brown, Department for Health, University of Bath, Claverton Down, Bath BA2 3AY, United Kingdom. Telephone: 4412-2538-4323. E-mail: D.J.Brown@bath.ac.uk

Abstract

Background. Psychologists are increasingly supporting the quest for performance enhancement in sport and there is a need to evaluate the evidence base underpinning their work.

Objectives. To synthesize the most rigorous available research that has evaluated psychological, social, and psychosocial interventions with sport performers on variables relating to their athletic performance, and to address some of the perplexing issues in the sport psychology intervention literature (e.g., do interventions have a lasting effect on sport performance?).

Methods. Randomized controlled trials were identified through electronic databases, hand-searching volumes of pertinent journals, scrutinizing reference lists of previous reviews, and contacting experts in the evaluation of interventions in this field. Included studies were required to evaluate the effects of psychological, social, or psychosocial interventions on sport performance in athletes when compared to a no-treatment or placebo-controlled treatment comparison group. A random effects meta-analysis calculating the standardized mean difference (Hedges' *g*), meta-regressions, and trim and fill analyses were conducted. Data were analyzed at posttest and follow-up (ranging from one to four weeks after the intervention finished) assessments.

Results. Psychological and psychosocial interventions were shown to enhance sport performance at posttest ($k = 35$, $n = 997$, Hedges' $g = 0.57$, 95% CI = 0.22, 0.92) and follow-up assessments ($k = 8$, $n = 189$, Hedges' $g = 1.16$, CI = 0.25, 2.08); no social interventions were included or evaluated. Larger effects were found for psychosocial interventions and there was some evidence that effects were greatest in coach-delivered interventions and in samples with a greater proportion of male participants.

Conclusions. Psychological and psychosocial interventions have a moderate positive effect on sport performance, and this effect may last at least a month following the end of the intervention. Future research would benefit from following guidelines for intervention reporting.

Key Points

- A meta-analysis was conducted to evaluate the effects of psychological, social, and psychosocial interventions on sport performance.
- High-quality studies show that psychological and psychosocial interventions can improve the sport performances of athletes.
- Improved reporting standards of intervention research will enable greater exploration of the differences in treatment effects.

1 Introduction

With psychologists increasingly supporting the quest for performance enhancement in sport [1-8], there is a need to evaluate the evidence base underpinning their work. Interventions developed and implemented by psychologists can be broadly categorized as psychological, social, or psychosocial in nature, which we describe as any actions or processes that alter functioning and/or performance through changes in an individual's thought and behavior, through social factors, or through a combination of both individual thought and behavior and social factors, respectively. To establish an evidence base for these treatments it is necessary for researchers to embark on a rigorous and iterative process of conceptualization, development, and testing conducted in both controlled clinical contexts and real-world settings [9-11]. The purpose of this review was to synthesize the most rigorous research that has evaluated the effects of psychological, social, or psychosocial interventions with sport performers on variables related to their athletic performance (viz., components of fitness, overall/competitive sport performance, and technical tasks).

Previous reviews of interventions in sport have typical focused solely on psychological techniques, and one of the earliest attempts to synthesize studies examining the effects of these interventions with athletes in competitive situations was conducted by Greenspan and Feltz [12]. Their seminal paper was the first review in this area to explicitly restrict inclusion to studies sampling athlete participants (i.e., those competing on a regular and organized basis) and reporting performance outcomes assessed in non-contrived competitive situations in the sport in which the participants regularly competed. Follow-up papers using similar criteria were published by Vealey [13] and Weinberg and Comar [14] with the former reviewing sport psychology intervention studies published between 1988 and 1991, and the latter reviewing studies published in 1992 and 1993. Collectively, these three reviews found that 45 studies had employed psychological interventions in competitive sport settings and that 38 (85%) had found positive performance effects, although this interpretation was largely based on the direction of observed effects and causality could only be inferred in 20 of these studies [14]. Since the turn of the century, reviews in sport have focused on studies that have evaluated interventions using single-subject [15], experimental [16], and single-case designs [17]. Regarding the review of the more rigorous experimental designs, Martin et al. [16] interpreted that 14 of the 15 interventions had a positive effect on sport performance. These findings should, however, be treated with a degree of caution because the search strategy was restricted since they, like the authors of the single-subject and single-case reviews, did not use electronic database searches and they excluded sports science and general psychology journals. Such incomplete paper retrieval can lead to the unsystematic identification of available literature, the introduction of publication biases, and to inaccurate treatment effects being concluded [18, 19].

As the number of intervention studies has increased, reviewers of sport research have tended to focus on specific types of psychological or social interventions, including mental practice [20-22], goal setting [23], team building [24, 25], self-talk [26, 27], and stress management [28]. These reviews all identified positive performance effects using either vote-counting procedures or effect size interpretation. A notable strength of this body of work is that it employed systematic procedures, typically including meta-analytic techniques. However, these reviews are limited by several methodological issues. First, the majority of the reviews in this area included studies that evaluated the effects of interventions on exercise-based or general motor tasks [21, 23]. Second, some of the reviews included studies that sampled nonathletes, typically students [20, 26]. The main limitation of these task and sampling issues is that they limit the generalizability and external validity of the findings for sport performance [11, 29]. Third, some of the reviews included studies with a wide range of research designs [24, 28]. The main limitation of this design issue is that it introduces uncertainty about whether the intervention caused the observed effects, thus compromising the inference of causality and internal validity [30, 31]. Thus, to overcome the limitations of previous reviews, only studies that had high internal (i.e., conducted with a randomized, controlled, experimental design) and external (i.e., sample sport performers and evaluate the intervention effect on variables relating to athletic performance) validity were included in the current review.

In addition to providing an overall evaluation on the effects of interventions on sport performance, the purpose of the current review was to address some of the perplexing issues in the sport psychology intervention literature. These issues include: whether interventions are effective for both sexes and across all competitive standards [3, 12, 16, 17], whether characteristics of the intervention provider influence treatment effectiveness [3, 14], whether effects vary based on the type of intervention and if single interventions are more or less effective than multicomponent interventions [3, 12, 13, 15-17], whether comparable effects are found across objective performance outcomes [3, 17], and whether interventions have lasting effects on sport performance [3, 12-14, 16, 17]. It is hoped that these supplementary analyses will provide practitioners with greater insight into who will benefit from interventions, which types of intervention are most effective, and how best to evaluate the impact of the treatments they provide.

2 Methods

2.1 Inclusion Criteria

Studies were required to meet the following inclusion criteria:

1) Evaluate the effects of an intervention that involved any action or process that used individual thought and behavior (e.g., hypnosis), social factors (e.g., teambuilding), or both individual thought and behavior and social factors (e.g., coach providing performance feedback) to alter functioning and/or performance.

2) Examine performance on an outcome that was either (a) a technical task required by athletes in their sport (e.g., golf putting), (b) a component of fitness required by athletes in their sport (e.g., strength), or (c) an overall performance or competition outcome (e.g., competition shooting performance).
3) Sample participants of any age who were competing in sport at local, regional, national, or international levels.
4) Constitute a parallel or cross-over randomized experimental design, with a no-treatment control or placebo-treatment comparison group that involved contact with the researcher but without an intervention being delivered[1]. Studies with a parallel randomized experimental design have two or more distinct conditions and participants only experience one condition; these have also previously been described as pretest-posttest control group designs and posttest-only control group designs [32]. Cross-over trials are those where participants receive all treatments in a random order.
5) Report sufficient statistical data to calculate effect sizes (i.e., sample sizes, means, and standard deviation for the groups, or information from other statistical tests that allowed for the computation of effect sizes with less exact estimation procedures such as an *F* test). If a study design was cluster-randomized (e.g., randomly allocated by team or training group), rather than individual participant-randomized, it was also necessary for the intracluster correlation coefficient to be reported.
6) Be available as a full-text in the English language or as a full-text translation prior to the end of the data collection period (viz. March 2015).

2.2 Search Strategies

To ensure comprehensive coverage of the available literature, four search strategies

[1] This inclusion criterion was established because sole treatments or comparisons between active treatments lack a meaningful comparison, particularly in view of the aforementioned limitations of previous reviews in this area of sport psychology research.

were used to retrieve relevant papers. First, electronic searches were conducted in on-line databases (viz., Applied Social Sciences Index and Abstracts, Biological Sciences, Physical Education Index, PsycARTICLES, PsycINFO, SCOPUS, SPORTDiscus, and Web of Science). A collection of search terms was used to encompass the psychological, social or psychosocial nature of treatments; the use of an athletic sample; a performance outcome; and an appropriate study design (the full search string can be viewed in Electronic Supplementary Material Appendix S1). The second search strategy involved hand-searching volumes of the pertinent journals[2]. In the third search strategy, the reference lists of previous meta-analytic, systematic, and narrative reviews papers were scrutinized for any unidentified studies that may have been relevant (see Electronic Supplementary Material Appendix S1 for a list of reviews). For the fourth search strategy, 27 researchers who were noted experts in the evaluation of psychological, social, or psychosocial interventions on sport performance[3] were contacted to retrieve any published or unpublished studies that they were aware of. The papers retrieved from the search strategies were evaluated by title, abstract, and full text. At each stage of the evaluation, studies were excluded from the winnowing process if the inclusion criteria were not satisfied.

The first author was responsible for screening titles, abstracts, and full text articles and for determining study eligibility. The second author reviewed a random sample (10%) of the papers at each stage of the winnowing process to ascertain if any studies had been erroneously included or excluded. Inter-coder reliability was calculated by dividing the

[2] Hand-searched journals include: *European Journal of Sport Science (2001 – 2015); International Journal of Applied Sport Sciences (2000 – 2015); International Journal of Sport and Exercise Psychology (2003 – 2015); International Journal of Sport Psychology (1970 – 2015); Journal of Applied Sport Psychology (1989 – 2015); Journal of Clinical Sport Psychology (2007 – 2015); Journal of Sport Psychology (1979 – 1987) and Journal of Sport and Exercise Psychology (1988 – 2015); Journal of Sport Behavior (1978 – 2015); Journal of Sport Psychology in Action (2010 – 2015); Journal of Sports Sciences (1983 – 2015); Psychology of Sport and Exercise (2000 – 2015); Research Quarterly in Sport and Exercise (2000 – 2015); Scandinavian Journal of Medicine and Science in Sports (1991 – 2015); Sport, Exercise, and Performance Psychology (2012 – 2015);* and *The Sport Psychologist (1984 – 2015).*

[3] Researchers were selected based on having published four or more sport-related intervention studies.

number of agreed studies by the total number of studies reviewed by both authors. For example, the authors agreed on the eligibility of 141 of 147 studies at the abstract stage (95.9% agreement). Overall, the percentage of agreement ranged from 92.7% to 96.1%. Any disagreements between the authors were resolved through discussion and debate until a consensus was reached. Where insufficient information was available to warrant study exclusion during the title and abstract stages of the evaluation, studies were retained in the sample. If information necessary for determining study eligibility was not reported at the full text stage of the evaluation, the study authors were contacted via email and requested to provide clarification. Detailed notes were recorded outlining the reasons for study inclusion/exclusion and the number of studies included and excluded at each stage. A visual summary of the study selection process is presented in Figure 1.

2.3 Study Coding Procedures

Studies that were suitable for inclusion in the review were coded independently by the authors using a standardized form for study characteristics and intervention characteristics. Any discrepancies in the coding process were resolved through discussion until an agreement was reached. Inter-coder agreement was 96.9%. Data were extracted for study design (i.e., parallel or cross-over design, matched or unmatched participants), participant characteristics (age, sex, competitive standard, matching variables (where appropriate), total sample size, type of sport), performance outcome details (i.e., component of fitness, overall performance or competition outcome, technical task), intervention characteristics (provider, setting, type), and descriptive information pertaining treatment integrity [33-35]. Regarding the latter, treatment/intervention integrity is akin to program integrity as described by Dane and Schneider [33] and has five aspects: adherence (the extent to which intervention components were delivered as prescribed), exposure (number, length and frequency of implementation of intervention components), quality of delivery (qualitative aspects of intervention delivery that are not directly related to implementation, e.g., training of implementers), participant responsiveness (measures of participant response to the intervention, e.g., satisfaction), and program differentiation (safeguard checks against the diffusion of treatments).

In addition to study and intervention characteristics, studies were also coded for study quality using the risk of bias method which involves the assessment of the impact of systematic error on the results and conclusions of a study [36, 37]. The Cochrane Collaboration's tool for assessing risk of bias [36] was used to assess the sources of bias which included selection bias, performance bias, attrition bias, detection bias, reporting bias, and other biases. This strategy involved writing descriptive accounts of a paper's method[4] (e.g., whether they used blinding of participants) and judging the risk of bias as low, high, or unclear. Additional considerations for studies that included multiple intervention groups were whether data were presented for each of the groups to which participants were randomized and whether the study was free of suggestion of selective reporting of comparisons. A detailed description of risk of bias variables and criteria for judgment is provided in Electronic Supplementary Material Appendix S2.

2.4 Statistical Procedures

Effect size estimates were computed for each study at posttest and, where data were provided, any follow-up assessment. To ensure that the parallel-groups and cross-over studies could be included within the same analysis, effect sizes in cross-over trials were computed using data from the initial treatment condition. This approach was also used to reduce the potential variability in risk of bias across, and the uncertain compatibility between, study designs [38].

Three steps were followed to calculate the effect size estimates and were conducted

[4] Descriptive accounts are available from the first author on request.

using the Comprehensive Meta-Analysis software package Version 3 (CMA) [39]. First, data were extracted from studies and imputed into the computer program. Second, standardized mean difference values were computed using one of four sets of formulae (see Electronic Supplementary Material Appendix S3) depending on the information available within the paper and irrespective of the initial study design. The first set of formulae (equations 1-4) calculated the standardized mean difference based on the means, standard deviations, and number of participants for each group. The difference in means was standardized by the standard deviation at the second time-point (i.e., posttest, follow-up). The second set of formulae (equations 5-9) calculated the standardized mean difference using data from independent groups at one time-point and the option for pooled variance. The third set of formulae (equations 1, 10-14) were used when studies provided within groups change scores for *t*, and the mean difference was standardized by the standard deviation of the change score. The fourth set of formulae (equations 1, 12-14) were used when study authors reported the within group change means and standard deviations, and the difference in means was standardized using the change score standard deviation. Where studies contained multiple performance outcomes, the mean effect on the outcomes was computed to create a single, averaged effect for the intervention. Furthermore, for studies which included different variations of the same type of psychological, social, or psychosocial intervention (e.g., instructional self-talk, motivational self-talk), an averaged effect was calculated. Hence, to ensure that the independent samples assumption [40, 41] was not violated, analysis was conducted on aggregated effect sizes at a study level. The third step in the effect size computation involved converting the standardized mean difference scores into Hedges' (adjusted) *g* values [42] (hereafter Hedges' *g*), by multipling the standardized mean difference with a correction factor, *J* (equations 15-18). Hedges' *g* was selected as the effect size measure because it accounts for variation in sample size and sample variance [43]. The magnitude of these effect sizes can be interpreted as small (0.2), moderate (0.5), and large (0.8) using Cohen's [44] anchors.

A random-effects (method of moments) computational model with Knapp-Hartung [45] modification was used to calculate the mean intervention effects. A random-effects model assumes that variability exists between studies and in the underlying effect for each study [46], and was thus selected to control for the systematic differences between studies (e.g., different practitioners). When using a random-effects model, there are three approaches that can be used to estimate the between-study variance (T^2): method of moments (MM), unrestricted maximum likelihood (ML), and restricted maximum likelihood (REML). The MM method was selected in the current study because it does not make any assumptions about the distribution of the random effects and thus offers a more robust analysis than the other two approaches [47]. The Knapp-Hartung [45] method was used to calculate the standard error for the random effects model. This approach calculates error based on the *t*-distribution (rather than the *Z*-distribution) to account for estimating the dispersion between studies as well as the variation within studies; the *Z*-distribution only accounts for the within study error [47]. The 95 percent confidence interval (95% CI) was calculated for each effect and an effect was deemed significant if the 95% CI did not include zero [48]. Furthermore, following the recommendation of Borenstein [49] to interpret the effect size in context, the mean intervention effect was compared to the effects reported in previous reviews of psychological, social, and psychosocial interventions in sport.

The data were checked for clinical and statistical heterogeneity. First, to protect against the potential impact of the variety in interventions and outcome variables between studies, each study was carefully re-examined against the inclusion criteria. Second, to assess statistical heterogeneity, five statistics were used: Q, p, tau squared (T^2), tau (T), and I^2. The Q statistic is a measure of weighted squared deviations and quantifies the total variance in a meta-analysis. The significance (*p*-value) of Q provides evidence that the true effects between studies vary. T^2 and T are computed from Q and provide measures of the variance and standard deviation of true effects, respectively. Also computed from Q is the I^2 statistic which is used to determine the proportion of the observed variance that is real (i.e., that is explained by between-study differences). T^2 and I^2 provide measures of the magnitude of heterogeneity and their 95% CIs provide support for whether any apparent heterogeneity is genuine. For example, if the lower limit of I^2 exceeds zero, then I^2 should be statistically significant and the between-study differences real. T was used to calculate the predictive intervals (PI) of the mean effect and enabled interpretation of the dispersion of true effects. Statistical heterogeneity was also examined graphically using a Galbraith plot [50] which enabled the identification of outliers.

A series of sensitivity analyses were then conducted to examine how robust the mean effect was to the influence of potential outlier studies, missing data, and the threat of publication bias. To investigate the effects of potential outliers, analyses were conducted with outlier studies included and excluded, and the results were compared to establish whether the conclusions drawn were substantially different. In relation to missing data, all studies were coded for incomplete or selective outcome reporting using the aforementioned Cochrane Collaboration risk of bias tool [36]. If the authors failed to provide data on some, or all, of the expected outcomes, or reported data incompletely so that they could not be entered into a meta-analysis, studies were deemed to be at a high risk of reporting bias [37]. To assess the impact of this missing data on the mean effect, a separate analysis was conducted with high-risk studies excluded. To determine how robust the effect size estimate was to the threat of publication bias, Duval and Tweedie's [51] trim and fill procedure was used. Displayed graphically on a funnel plot, this procedure involves systematic removal (i.e., trimming) of extreme small studies on the positive side of the plot until the plot is symmetric about a new effect size. The algorithm then adds (i.e., filling) the studies back to the plot with imputed mirror image studies, thus creating an unbiased estimate of the effect size. An effect size estimate is likely to be robust to the effect of publication bias if the unbiased estimate is not > 0.05 above or below the original value.

To assess the relationships between potential covariates (moderators) and the effect size estimate, a series of exploratory post-hoc analyses were conducted using the meta-regression analysis module within the CMA software [39]. The decision to use meta-regression techniques instead of more common subgroup procedures was made to enable the consistent use of the Knapp-Hartung [45] modification across all analyses. Through these exploratory analyses, the effects of six potential moderator variables were considered: sex (proportion of male participants in the sample), participants' competitive standard (local, regional, national, international), intervention provider (e.g., coach, practitioner, researcher), type of intervention (psychological, psychosocial, social), whether the intervention was uni- or multi-modal, and the type of outcome characteristic (component of fitness, overall performance or competitive outcomes, technical task). In view of the need to maintain an appropriately large ratio of studies to covariates and for studies to provide data on the treatment effect, variance, and covariate values [52], separate meta-regressions were conducted on each of the moderators. It is acknowledged, however, that this may have hidden potential confounding effects between covariates and thus the correlation matrices between moderating variables were studied to ensure that the variables were unrelated. The test of model statistic (F-ratio) was used to assess whether the covariate (continuous or categorical) coefficients were significantly different from zero and if the variable was likely to have a moderation effect. Changes in the goodness of fit statistics (T^2, Q) between the original and covariate models were used to assess whether the covariates accounted for any of the unexplained variance in the model.

3 Results

3.1 Search Results and Study Characteristics

Systematic retrieval identified 3,174 potentially relevant papers. Searches of electronic databases returned 2,104 papers (1,366 after duplicates removed), 550 (496) were retrieved from hand searching journal volumes, 514 (450) from the reference lists of previous narrative, systematic, and meta-analytic reviews, and 6 (6) from contacting potential study authors. Retrieved papers included published journal articles, books, book sections, conference proceedings, unpublished manuscripts, and theses. Figure 1 depicts the winnowing process, provides explanations for the exclusion of papers, and shows that 35 full-texts (35 studies) met all of the inclusion criteria; all of these were published journal articles. Eligibility could not be determined for 21 papers because insufficient information was provided (e.g., a lack of information about the allocation of participants to condition). Furthermore, 41 studies were identified that met the majority of the review criteria but failed to report or provide the descriptive statistics necessary for the computation of an effect size and inclusion in a random-effects meta-analysis. After examining the nature of the interventions included within the studies, one study [53] was excluded because the ego depletion intervention delivered was designed to negatively impact performance and was therefore considered clinically heterogeneous.

The descriptive information for the studies eligible for inclusion in the review are provided in Table 1 (and more detailed study specific information can be found in the three additional tables (Tables S1-S3) presented in the Electronic Supplementary Material Appendix S4). Of the 35 studies suitable, 34 were categorized as parallel individual randomized controlled trials (RCT) and one as cross-over individual randomized controlled trials (RXCT). No identified cluster randomized trials were suitable for inclusion. Matching of participants prior to random allocation was used in six of the 35 studies. In terms of the risk of bias and the impact of systematic error, the majority of studies were classified at low risk for random sequence generation (91.4%), blinding of outcome assessment (88.6%), incomplete outcome data (91.4%), and selective outcome reporting (91.4%). However, 94.3% of studies failed to report sufficient information to determine the risk of bias pertaining to allocation concealment and 100% of studies were classified as either high or unknown risk of bias for the effect of blinding of participants and personnel.

Turning to participant characteristics, the largest number of studies (48.6%) recruited a male only sample and the age of participants included in the studies ranged from 14 years to over 62 years. Sport performers' competitive standard ranged through local (42.9%), regional (5.7%), national (17.1%), and international (2.9%) levels and performers were recruited from 16 different sports (e.g., basketball, golf, soccer). In total, 58 interventions were delivered across the 35 studies (see Table 1 for a summary of intervention types and Electronic Supplementary Material Appendix S4 Table S2 for a detailed description of each intervention delivered), including 46 psychological interventions and 12 psychosocial interventions; no social interventions were identified. The most frequent types of interventions were perceptual training (i.e., techniques designed to alter performers' visual strategies; 15.5%) and multi-modal pre-performance routines (i.e., combinations of two or more techniques to be implemented prior to a performance commencing; 13.8%) and, within the various types of intervention, different subtypes were identified. For example, feedback-based interventions included biofeedback and visual feedback. A combination of uni- (60.3%) and multi-modal (39.7%) interventions was delivered, with the main provider being a researcher (62.1%). Nineteen of the interventions were designed or delivered individually and the remaining 39 were provided in a group setting. Each of the three different types of performance outcome were assessed across the studies with performance assessed most frequently using technical tasks (56.9%) such as golf putting and basketball free-throw shooting. The components of fitness (5.6%) assessed included speed and strength, and the overall performance or competitive outcomes (37.5%) assessed included the number of correct in-match decisions, judges' score of a karate performance, and competitive shooting performance.

3.2 Meta-Analytic Results

Prior to calculating a mean effect, studies were checked for clinical heterogeneity. In addition to the one excluded study described in section 3.1, two further interventions were deemed clinically heterogeneous to the other treatments because they were designed to negatively disrupt or impair performance, and were therefore removed from subsequent analyses. Specifically, these were an inaccurate biofeedback intervention [54] and a dual attentional focus manipulation [55]. Effect size calculations were computed for the remaining comparisons at posttest and follow-up assessments (see Table 2).

Thirty-five studies [54-88] ($n = 997$) were then combined in a random-effects (MM) model with Knapp-Hartung [45] modification. The standardized mean difference (Hedges' *g*) was 0.57 with a 95% CI of 0.22 to 0.92 and a PI of -0.68 to 1.82 (see Table 3, model 1). This moderate mean effect was larger than the majority of effects observed in previous reviews of interventions in sport (e.g., 0.26, Driskell et al., [20]; 0.34, Kyllo & Landers, [23]; 0.48, Hatzigeorgiadis et al., [26]). However, the 95% CI was wide suggesting that the observed mean effect estimate lacked precision and substantial uncertainty existed. Furthermore, the PI was very large suggesting that an intervention in a future study may have a positive, null, or negative effect. Interpretation of the heterogeneity statistics suggested that differences existed in the true effects observed in the studies. To elaborate, the magnitude of heterogeneity was illustrated by the variance predicted in true effect sizes ($T^2 = 0.35$) and by the high proportion of variance explained by between-study differences ($I^2 = 69.91\%$). Additionally, the p-value (< 0.001) for $Q = 113.00$ with $df = 34$, and the 95% CIs for T^2 [0.20, 0.55] and I^2 [57.70%, 78.60%] supported the presence of statistical heterogeneity.

To determine how robust the pooled effect was to the influence of potential outlier studies, systematic error resulting from reporting bias, and publication bias, a series of sensitivity analyses were conducted. The first analysis identified two studies [59, 68] that may be considered outliers. More specifically, the Hedges' *g* values for Caserta et al. [59] (n = 18, Hedges' g = 6.94, 95% CI 4.43, 9.45) and Lorains et al. [68] (n = 30, Hedges' g = 6.00, 95% CI 4.32, 7.69) were substantially larger than the effect sizes computed for the other studies (range -0.97 to 1.35), which was also apparent visually in Figure 2. When these studies were removed from the combined analysis (see Table 3, model 2, n = 950), the standardized mean difference (Hedges' *g*) was 0.43 with a 95% CI of 0.26 to 0.59 and a PI of -0.10 to 0.95. Therefore, although greater precision in the estimate was apparent in model 2 when compared to model 1, the effect size estimates remained comparable and the conclusions drawn from both models were not substantially different. However, the heterogeneity reduced in model 2 and the amount of total variance was non-significant (Q = 45.48, df = 32, p-value = 0.058), suggesting that some of the true difference in the study effect sizes in model 1 may have been due to the two outlier studies. Second, to assess the potential effect of reporting bias, an analysis was conducted with studies excluded if they were deemed to be at high risk (i.e., studies that had incomplete reporting or selective reporting of statistical results). Three studies [61, 66, 88] met this exclusion criterion. The results obtained from the revised analysis (see Table 3, model 3, n = 915) were comparable to those found in model 1, which therefore suggested that the original analysis was not influenced by selective outcome reporting. The final sensitivity analysis involved the assessment of publication bias on the estimate from model 1. To assess publication bias, Duval and Tweedie's [51] trim and fill procedure was implemented and resulted in zero additional studies being imputed (see Figure 3, k = 35, g = 0.57). This analysis therefore suggested that the summary mean effect was also robust to the threat of publication bias.

Eight studies included follow-up assessments to determine the enduring effect of the interventions (see Table 2). The reported length of time between pretest and follow-up assessment ranged from two to six weeks. In order to provide a meaningful comparison for these effects it was necessary to determine the combined effect of these eight studies at posttest. Computed using a random-effects (MM) model with Knapp-Hartung [45] modification, the standardized mean difference (Hedges' *g*) for the eight studies (n = 189) at posttest was 1.02 with a 95% CI of -0.37 to 2.41 and a PI of -1.72 to 3.77 (see Table 3, model 4). The standardized mean difference (Hedges' *g*) for the eight studies (n = 189) at follow-up was 1.16 with a 95% CI of 0.25 to 2.08 and a PI of -1.12 to 3.45 (see Table 3, model 5). These findings showed that the interventions had a large positive effect on sport performance when assessed at follow-up; however, these effects varied considerably and readers should be cautious when interpreting the magnitude of the effect. Similar to the posttest analysis of the heterogeneity for these studies, the follow-up heterogeneity statistics suggested that differences existed in the true effects observed in the studies (Q = 31.61, df =7, p < 0.001, T = 0.85). Sensitivity analysis suggested that the pooled effect may have been influenced by an outlier study [68] (see Table 3, model 6 and Figure 4), but was robust to the threat of reporting bias (see Table 3, model 7) and publication bias (see Figure 5).

3.3 Moderator Analysis Results

Exploratory meta-regression procedures were conducted to investigate the between-study variance apparent in the posttest model. Specifically, analyses were run to examine the potential moderating effects of six variables: two participant characteristics (competitive standard, sex), three intervention characteristics (provider, single or multiple component, type), and one outcome characteristic (type of performance assessment). The test of model and goodness of fit statistics for these six analyses are shown in Table 4 and the estimated mean effects across the participant, intervention, and outcome characteristics are displayed graphically in figures 6-8, respectively. It is important to note that these moderator analyses should be interpreted cautiously given their exploratory nature and the reduced statistical power inherent in meta-regression procedures.

3.3.1 Participant Characteristics

The first meta-regression assessed the effect of participant sex (% male participants) and excluded studies [57, 74, 79] which did not report sex; the standardized mean difference (Hedges' *g*) for the remaining 32 studies ($n = 950$) in the intercept model was 0.60 with a 95% CI of 0.24 to 0.97 and a PI of -0.64 to 1.85. The goodness of fit statistics for this revised intercept model (see Table 4, model 8, $T^2 = 0.34$, $Q = 104.77$, $df = 31$, $p < 0.001$) were comparable to those found in the full posttest model (see Table 4, model 1, $T^2 = 0.35$, $Q = 113.00$, $df = 34$, $p < 0.001$). A covariate model (see Table 4, model 9) containing sex was then assessed in comparison to the revised intercept model. The test of model statistics for the covariate model ($F(1, 30) = 0.86$, $p = 0.360$) suggested that the proportion of males in the sample was not likely to influence the effect size estimate. Furthermore, the goodness of fit statistics suggested that a significant level of heterogeneity remained in the model following the introduction of the covariate ($Q_{\text{model 9}} = 98.37$, $df = 30$, $p < 0.001$). However, the variance of true effect sizes (T^2) was marginally reduced in the covariate model when compared to the intercept model (0.33 and 0.34, respectively) with $R^2 = 0.04$. The proportion of males in the sample therefore accounted for 4% of the unexplained variance in the intercept model.

In the second analysis examining performers' competitive standard, studies that included mixed competitive standards [54, 55, 59, 65, 67, 71, 72, 76, 79-81] were excluded; the standardized mean difference (Hedges' *g*) for the remaining 24 studies ($n = 783$) was 0.56 with a 95% CI of 0.21 to 0.90 and a PI of -0.49 to 1.60. Table 4 displays the goodness of fit test statistics and suggests that the variance in true effects between the studies included in this analysis (model 10, $T^2 = 0.23$) was smaller than that found in the full posttest model (model 1, $T^2 = 0.35$). A meta-regression was then conducted with participants' competitive standard entered as covariate (see Table 4, model 11). The test of model statistics ($F(3, 20) = 0.33$, $p = 0.802$) suggested that the competitive standard was unlikely to influence the effect size estimate. Furthermore, the goodness of fit statistics from this analysis suggested that the covariate model failed to explain any additional variance within the data with between-study variance greater in the covariate model ($T^2 = 0.26$, $Q_{model\ 11} = 60.73$, $df = 20$, $p < 0.001$) compared to the intercept model ($T^2 = 0.23$, $Q_{model\ 10} = 63.52$, $df = 23$, $p < 0.001$). Thus, the intervention effect did not differ across competitive standards.

3.3.2 Intervention Characteristics

Three intervention characteristics were assessed: provider, single or multiple components, and type. One study involving assessments of interventions from different providers [60] was excluded from the meta-regression on intervention provider; the standardized mean difference (Hedges' *g*) for the remaining 34 studies ($n = 937$) was 0.59 (95% CI = 0.23, 0.95; PI = -0.69, 1.87). The goodness of fit statistics for this model (see Table 4, model 12) showed that T^2 was comparable in this intercept model (0.36) to the full model. A covariate model (see Table 4, model 13) containing intervention provider was then assessed in comparison to the revised intercept model. The test of model statistics for the covariate model ($F(4, 29) = 1.45$, $p = 0.244$) suggested that the provider of the intervention was not likely to influence the effect size estimate. Furthermore, the goodness of fit statistics suggested that a significant level of heterogeneity remained in the model following the introduction of the covariate ($Q_{\text{model 13}} = 91.99$, $df = 29$, $p < 0.001$). However, the variance of true effect sizes (T^2) was marginally reduced in the covariate model when compared to the intercept model (0.34 and 0.36, respectively) with $R^2 = 0.05$. The intervention provider therefore accounted for 5% of the unexplained variance in the intercept model.

A meta-regression was conducted to establish if whether the intervention included single or multiple components influenced the intervention effect. The intercept model for this meta-regression (k = 29, n = 840, studies excluded [60, 73, 74, 80, 81, 87], see Table 4, model 14) showed an increase in T^2 (0.41) compared to the full model. Furthermore, when intervention components were entered into the model as a covariate (see Table 4, model 15), the level of between-study variance increased further ($T^2 = 0.42$) suggesting that the covariate did not explain any of the variance in the observed effect and that intervention with single or multiple components were comparable.

Studies involving assessments of different types of intervention (i.e., psychological and psychosocial) were excluded from the meta-regression on intervention type [60, 74, 80, 81]; the standardized mean difference (Hedges' g) for the remaining 31 studies (n = 880) was 0.59 (95% CI = 0.19, 0.99; PI = -0.75, 1.93). The goodness of fit statistics for this model (see Table 4, model 16) showed that T^2 was marginally greater in this intercept model (0.39) compared to the full model. A covariate model (see Table 4, model 17) containing type of intervention was assessed in comparison to the revised intercept model. The test of model statistics for the covariate model ($F(1, 29) = 9.08$, $p = 0.005$) suggested that the type of intervention was likely to influence the effect size estimate. Furthermore, the variance of true effect sizes (T^2) was reduced in the covariate model when compared to the intercept model (0.32 and 0.39, respectively) with $R^2 = 0.20$. The intervention type therefore accounted for 20% of the unexplained variance in the intercept model. However, the goodness of fit statistics suggested that a significant level of heterogeneity remained in the covariate model ($Q_{\text{model 19}} = 90.62$, $df = 29$, $p < 0.001$).

3.3.3 Outcome Characteristics

A meta-regression assessed the potential moderating role of type of performance outcome on intervention effect. One study involving different types of performance outcomes [69] was excluded from this analysis; the standardized mean difference (Hedges' *g*) for the remaining 34 studies (n = 980) was 0.58 with a 95% CI of 0.23 to 0.94 and a PI of -0.68 to 1.85. The goodness of fit statistics for this intercept model (see Table 4, model 18) demonstrated that the studies included in this meta-regression were comparable to those used in the full posttest model (model 1). The introduction of type of performance outcome as a covariate in this analysis did not reduce the level of unexplained variance in the model ($T^2_{model\ 19} = 0.38$; $T^2_{model\ 18} = 0.36$). This finding was supported by the test of model statistics ($F(2,31) = 0.15$, $p = 0.862$), which showed that the coefficients were not significantly different from zero. Thus, the intervention effect did not vary across the three types of performance outcome (i.e., components of fitness, overall/competition performance, technical task).

3.3.4 Combined Model

The last meta-regression assessed the collective impact of the moderators that had been found to explain some of the unexplained variance in intervention effects when considered independently. Correlations between participant sex, intervention provider, and type of intervention[5] suggested that relationships were apparent between the variables. The three variables were then included simultaneously in a meta-regression with studies excluded if they failed to provide information on participant sex, included multiple types of intervention provider, or included different types of interventions. This resulted in six studies being excluded [57, 60, 74, 79-81]; the standardized mean difference (Hedges' *g*) for the remaining 29 studies (n = 850) was 0.62 (95% CI = 0.21, 1.02; PI = -0.71, 1.94). The goodness of fit statistics for this model (see Table 4, model 20) showed that T^2 was comparable in this intercept model (0.38) to the full model. The test of model statistics for the covariate model (see Table 4, model 21; $F(6, 22) = 2.96$, $p = 0.028$) suggested that

[5] Correlation matrix available from the first author on request.

participant sex, the provider of the intervention, and the intervention type were likely to influence the effect size estimate. Furthermore, the variance of true effect sizes (T^2) was substantially reduced in the covariate model when compared to the intercept model (0.27 and 0.38, respectively) with $R^2 = 0.28$. The combined effect of the three variables therefore accounted for 28% of the unexplained variance in the intercept model. However, the goodness of fit statistics suggested that a significant level of heterogeneity remained in the model following the introduction of the covariates ($Q_{\text{model 21}} = 62.49$, $df = 22$, $p < 0.001$) and that other factors may still exist.

4 Discussion

4.1 Discussion of Findings

The purposes of this study were to systematically review research that evaluated the effects of psychological, social, or psychosocial interventions with sport performers on variables relating to their athletic performance, and to address some of the perplexing issues in the sport psychology intervention literature (e.g., do interventions have a lasting effect on sport performance?). From an initial sample of 3174 potentially relevant papers, 35 met the inclusion criteria for the review. The findings from the posttest analysis showed that interventions had a moderate positive effect on sport performance; however, this conclusion was restricted to psychological and psychosocial techniques because no suitable studies evaluating social interventions were identified. Notwithstanding this overall positive finding, it is important to highlight that the large confidence interval indicated a lack of precision in the mean effect and that the prediction interval displayed a high level of dispersion in effect sizes. The effect found in the current review was generally larger than the positive conclusions drawn in previous general reviews [12, 14, 16] and reviews that have focused on specific types of psychological or social interventions. To elaborate, the effect size estimate was 0.57 compared to 0.26, 0.48 and 0.68 for mental practice [20-22], 0.34 for goal setting

[23], 0.43 for team building [24], and 0.48 for self-talk [26]. The more rigorous inclusion criteria and sampling of studies with high internal and external validity in this review, supported by a convergence of other published evidence, indicate that psychological and psychosocial interventions can enhance sport performance.

The findings from the follow-up analysis showed that interventions had an overall large positive effect on sport performance at least a month after the intervention had finished. Indeed, for studies included in this analysis, the effect of interventions became statistically significant when assessed at follow-up, compared to a non-significant finding when measured immediately following the delivery of the intervention. The lasting effect of interventions was possibly due to the participants either receiving a continued residual benefit from the intervention or continuing (whole or part) implementation of the intervention, or that the intervention group's decrement in performance after the intervention finished was less than that observed in the control group. Another explanation might be that the interventions have been delivered to alter intermediary psychological variables (e.g., self-efficacy) that act as mechanisms through which treatments influence performance [56], and it may be the case that a certain length of time is necessary for these changes to manifest and result in performance enhancement. Although all of these explanations suggest a favorable learning effect, the findings of the follow-up analysis should be interpreted cautiously for a number of reasons. First, only eight studies reported follow-up assessments, which mean that nearly 80 percent of the studies reviewed failed to examine the enduring effect of interventions. Second, the observed effect was imprecise and anticipated to range substantially. Third, the predictive interval suggests that the intervention effect in future studies may range from highly negative to highly positive.

The results suggested that a high level of imprecision existed in the posttest mean effect and substantial variability was apparent between studies. To explore these findings and to address the questions posed in the introduction, a series of sensitivity and moderator analyses were conducted. Sensitivity analyses showed that the mean effect was at greatest risk to outlier studies (in comparison to reporting and publication biases), with a smaller, more precise mean effect and a considerable reduction in heterogeneity found for the model with outliers removed; however, the differences between the models were not substantially different. Moderator analyses were conducted to explore the high level of between-study difference and to examine the relationships between participant, intervention, and outcome variables and the treatment effect. Specifically, these explored whether participant sex, participant competitive standard, intervention provider, multi-component nature of interventions, intervention type, and performance outcome characteristic reduced the variance in true effects and accounted for any of the unexplained variance in the model. Three of the moderators (participant sex, intervention provider, intervention type) were found to explain a portion of the variance when considered independently, and collectively accounted for 28% of the unexplained variance in the intervention effect. However, the level of heterogeneity remained significant suggesting that other factors not examined in the current review (e.g., psychological determinants, quality of therapeutic relationship), may provide further explanation for the effect of interventions on performance.

In-line with extant intervention research in sport psychology, the most abundant type of interventions included in this review were psychological in nature; that is, they were designed to alter functioning and/or performance through changes in an individual's thought and behavior. In total, 26 studies were identified that exclusively examined the effects of psychological interventions with the remaining five studies in this analysis testing psychosocial techniques. The lack of suitable studies evaluating the effects of social interventions highlights a notable gap in the literature needing to be addressed by future research. Figure 7 displays the estimated mean effects of psychological and psychosocial intervention types and shows that psychosocial interventions may be more effective than psychological interventions for enhancing sport performance. Based on the definition presented in section 1, psychosocial interventions attempt to bring about change through alterations in an individual's thought and behavior, and through social factors, and it may be the case that the added social component in these techniques supports and reinforces the changes occurring in the individual. More specifically, four of the five psychosocial interventions supported psychological components (e.g., personal performance footage) with guidance and questioning from a coach or researcher and, thus, the two separate components may interact to facilitate performer learning and accelerate the change needed to enhance performance. The remaining psychosocial intervention attempted to alter a performer's decision-making through the provision of concurrent performance feedback. When interpreting the findings from the intervention type analysis, it is important to consider that the moderation effect may have been amplified by the presence of two potential outlier studies [59, 68] in the psychosocial intervention category.

The finding that participant sex accounted for some of the unexplained variance in the intervention effect warrants further consideration. This review found that interventions were marginally more effective for samples containing a greater proportion of males. In addition, it was observed that male athletes were included in the majority of the participant samples (at least 32 of 35), with females sampled in approximately half (at most 18 of 35). From these two observations, it may be speculated that existing interventions have been developed, tested, and refined on male athletes more readily than on female athletes, and that a sex bias has resulted in the development of interventions that are effective for males, but not necessary females. One potential avenue for exploration relevant to sex would be whether interventions delivered in sex concordant dyads (i.e., same sex practitioner and performer) were more effective than those delivered in non-concordant dyads. Interestingly, previous research within the healthcare domain has found that the sex of the patient and practitioner affects the level of communication, the quality of the therapeutic relationship, and the level of treatment satisfaction, with results generally most favorable in sex concordant dyads [89, 90].

Although an examination of the effect of intervention provider sex was not conducted in the current review, a moderator analysis was run on the type of provider (i.e., coach, equipment, practitioner, researcher, self). The results for this analysis suggested that, despite intervention provider not representing a significant moderator, it did account for some of the variation in the observed effect. Perusal of Figure 7 shows that the coach and practitioner provider types introduced variability for this moderator with the estimated mean effects highest for coach delivered interventions and lowest for those delivered by qualified practitioners. The limited number of studies in these two categories and the inclusion of a potential outlier study in one of these precludes a reliable interpretation of this variability, but the former point raises the question why so few sport psychology interventions appear to have been developed and delivered by qualified professionals. The answer may lie in the typically poor standards of reporting apparent in the extant intervention literature. More specifically, we anticipate that some of the researchers in the included studies were also accredited practitioners; however, this information was not included in the published studies and, without it, the providers' qualifications cannot be assumed.

The remaining moderator analyses for participant competitive standard, intervention components, and type of performance outcome showed that these variables were unrelated to the intervention effect. Taken in turn, included studies recruited participants across a range of competitive standards and these samples suggested a movement away from a reliance on college athletes [12, 17]. Furthermore, the results provided initial evidence to suggest that athletic ability does not moderate intervention effects [3]. Numerous reviews of intervention research in sport [3, 12, 13, 15, 17] have queried whether single component interventions were more or less effective than multi-model interventions and have encouraged the analysis of the component parts of the latter. The moderator analysis in this review compared single and multiple part interventions and found no difference in the observed effects. However, results from the type of intervention analysis suggested that psychosocial interventions were more effective than psychological treatments, and it may therefore be the case that multiple part interventions are more effective, but only if they include both psychological and social components. The final moderator analysis examined whether the effect of interventions was consistent across different types of performance outcome [3, 17]. No moderation effect was found suggesting that intervention effects established on overall performance in competition or aspects of performance assessed in competition were similar to those found on technical tasks and components of fitness assessed away from a competition setting.

4.2 Applied Implications

The collective findings from this review have a number of implications for applied practice. First, the overall significant positive effect of interventions on sport performance provides a robust evidence base for the use of these techniques and affords credibility for the profession. Indeed, psychological and psychosocial interventions appear to have a substantial effect on performance and may therefore provide the critical, marginal gain often sought after in sport. Second, psychosocial interventions (i.e., those that included social factors alongside attempts to change thought and behavior) were found to be most effective, which suggests that supporting psychological techniques (e.g., goal setting, imagery) with an active social agent (e.g., coach), have a greater effect than simply providing the technique alone. Third, initial evidence appears to suggest that coaches are the most effective provider of interventions, potentially as a result of the performer and his or her coach possessing a more matured relationship and established rapport [91]. When delivering interventions, it may therefore be beneficial for practitioners to engage athletes' coaches to elicit greatest effects.

Fourth, the marginally greater effects for male performers compared to female performers may highlight one of the many small nuances that underpin the provider-performer relationship. We have discussed the potential influence of sex concordance in section 4.1; however, this may also extend to concordance in age, ethnicity, and sporting experience, for example. It would be beneficial for practitioners to have an awareness of these factors and how they could affect the interpersonal bond between the provider of the intervention and the athlete. Fifth, interventions were found to have comparable effects across performers' competitive standards and irrespective of the performance outcome used, which suggest that these techniques can have positive performance effects for a range of athletes and that any effects should be apparent in training and in competition.

4.3 Limitations and Suggestions for Future Research

Intervention review papers can be limited by variables at a study level and/or a review level [92]. Within this review, the main limiting factor at study level was the poor standard of reporting which resulted in 62 studies failing to meet the inclusion criteria for the review and included studies failing to provide descriptive information pertaining treatment integrity. To elaborate on the failure of studies to meet the inclusion criteria, 41 studies were excluded because they failed to report or study authors were subsequently unable to provide statistical data to calculate effect sizes estimates. This was particularly problematic for our review because we used a random effects model; if a fixed effects model had been used then advanced techniques for combining effect size estimates and vote-counts could have been implemented [93, 94]. An additional 21 studies were omitted because insufficient information was available to determine eligibility for inclusion. Turning to the issue of treatment integrity, we attempted to follow Dane and Schneider's [33] recommendations and, in so doing, provide readers with information pertaining to adherence, exposure, quality of delivery, participant responsiveness, and program differentiation for all of the interventions delivered in the included studies. However, many studies failed to provide information for one or more of these areas (see Electronic Supplementary Material Appendix S4 for details of each study). To prevent reporting shortcomings occurring in future intervention research, researchers are encouraged to prepare their work in accordance with established guidelines such as the CONSORT statement [95] or the American Psychological Association journal article reporting standards [96]. It is acknowledged that journal page restrictions make comprehensive reporting of the desired information for evaluating interventions difficult; however, to overcome this challenge, researchers should better exploit the expanding provision of on-line supplementary materials. Achieving higher standards of reporting for program integrity would enable the examination of variables pertinent to the delivery of the intervention that cannot currently be assessed (e.g., quality of the therapeutic relationship), and which may influence the observed intervention effect [97].

An additional study level limiting factor was risk to performance and selection biases. Susceptibility to performance bias was problematic owing to the unclear or high risk of bias from a lack of blinding of participants and researchers. However, the nature of psychological, psychosocial, and social interventions makes it very difficult, and in some circumstances impossible, to blind the participants to the intervention they are receiving and to blind the researchers to the intervention they are delivering and, therefore, although being classified as high or unclear risk of bias, this represents a somewhat severe critique. Uncertainty around selection bias was introduced as a result of the lack of information on the method of randomization used and, subsequently, how the allocation sequence was concealed. Although the latter criterion can be overcome through the adoption of the above suggested reporting standards, the former critique on performance bias would support the development of a study quality assessment tool specific to the demands of sport science or psychology research [3]. A meta-analysis may also be limited by biases arising through the search process. The decision was made in this review to retrieve both published and unpublished studies but the inaccessible nature of much of the unpublished literature meant that it was important to assess whether the overall intervention effects were robust to the threat of publication bias. Positively, both analyses were found to be robust to this threat.

The findings of a review are limited by the scope and retrieval of the studies included within it. To this end, the current review is limited to the examination of interventions delivered at an individual level because no cluster randomized trials at either the team or organizational level reported adequate statistics to be included. Furthermore, no studies were found that rigorously assessed the impact of social interventions on sport performance. More intervention research is therefore required to provide an evidence-base for team- and organizational-level interventions and, as this literature builds, it may also be of benefit for scientific inquiry to investigate whether intervention effects differ when delivered in a one-to-one context or as part of a group. An additional limitation of the included studies was the limited use of follow-up assessment. Only eight of the 35 studies meeting the inclusion criteria for the review provided data from a follow-up assessment. More studies are required that utilize follow-up assessments and that longitudinally track the intervention effect because this could help to identify the sources of variation, assess the fluctuations in the intervention effect over time, and determine the impact of the timing of an intervention. In relation to participant characteristics, there were only a small number of studies which sampled female only participants, international level participants, and athletes with a disability. Intervention researchers should attempt to sample more participants within these categories to increase the generalizability of the body of literature. One approach that could be used to achieve this greater breadth of sample would be to use study designs (e.g., single-case experimental designs) that enable the assessment of treatment effects across performer groups who may not be numerous enough to participate in a groups-research design (e.g., high achieving outliers).

It must be emphasized, however, that this approach should not compromise the drive for rigorous intervention research and neither will it prove successful without a coherent, systematic, and progressive approach.

Finally, the current review is limited to the study of the effects of interventions on sport performance. We acknowledge that interventions are often delivered to simultaneously target psychological, psychosocial, or social variables that may act as independent outcomes or as mechanisms for the performance effect; thus, it would be of benefit if future research examined the effect of interventions on these additional outcomes. Systematic scientific inquiry in this area, conducted with sport performers, has the potential to elicit a better understanding of how these interventions work and highlight if some techniques are effective as remedial treatments when performers wish to overcome a mental health or wellbeing issue in addition to achieving performance gains.

5 Conclusion

A meta-analysis was conducted to examine the effects of psychological, psychosocial, and social interventions on sport performance. Data from 35 randomized controlled trials were synthesized and results indicated that psychological and psychosocial interventions had a moderate performance effect, and that this positive effect may last up to a month following the end of the intervention. Furthermore, moderator analyses showed relationships between the intervention effect and intervention type, intervention provider, and performer sex. More specifically, intervention effects were greatest for psychosocial interventions, and marginally better when delivered by the coach and to male athletes. The findings from this review therefore provide an evidence base for the use of psychological and psychosocial techniques with sport performers and offer insight into the variables that may influence this effect. That said, the results suggested that a high level of imprecision existed in the posttest mean effect and that there was substantial heterogeneity between studies, which means that some level of caution is necessary when interpreting the findings. To improve understanding in this area, it is recommended that intervention research is published in accordance with reporting guidelines to ensure that greater detail on salient variables (e.g., quality of therapeutic relationship, randomization method) is provided.

Compliance with Ethical Standards

Funding

No sources of funding were used in the preparation of this review.

Conflict of Interests

Daniel J. Brown and David Fletcher declare that they have no conflicts of interest relevant to the content of this review.

Acknowledgements

The authors thank Thomas Curran for his advice about the statistical analysis and for his comments on an earlier version of this paper.

References

1. Gardner F, Moore Z. Clinical sport psychology. Champaign, IL: Human Kinetics; 2006.

2. Hackfort D, editor. Psycho-social issues and interventions in elite sport. Frankfurt, Germany: Lang; 1994.

3. McCormick A, Meijen C, Marcora S. Psychological determinants of whole-body endurance performance. Sports Med. 2015;45:997-1015. doi:10.1007/s40279-015-0319-6.

4. Meyers AW, Whelan JP, Murphy SM. Cognitive behavioral strategies in athletic performance enhancement. In: Hersen M, Eisler RM, Miller PM, editors. Progress in behavior modification. Pacific Grove, CA: Brooks/Cole; 1996. p. 137-64.

5. Murphy SM, editor. Sport psychology interventions. Champaign, IL: Human Kinetics; 1995.

6. Tod D, Edwards C, McGuigan M, et al. A systematic review of the effect of cognitive strategies on strength performance. Sports Med. 2015;45:1589-602. doi:10.1007/s40279-015-0356-1.

7. Vealey RS. Mental skills training in sports. In: Tenenbaum G, Eklund RC, editors. Handbook of sport psychology. New York City, NY: Wiley; 2007. p. 287-309.

8. Whelan JP, Meyers AW, Berman JS. Cognitive-behavior interventions for athletic performance enhancement. Annual Meeting of the American Psychological Association; New Orleans, LA1989.

9. Barkham M, Mellor-Clark J. Bridging evidence-based practice and practice-based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. Clin Psychol Psychother. 2003;10:319-27. doi:10.1002/cpp.379.

10. American Psychological Association Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. Am Psychol. 2006;61:271-85. doi:10.1037/0003-066X.61.4.271.

11. American Psychological Association. Criteria for evaluating treatment guidelines. Am Psychol. 2002;57:1052-9. doi:10.1037/0003-066X.57.12.1052.

12. Greenspan MJ, Feltz DL. Psychological interventions with athletes in competitive situations: A review. Sport Psychol. 1989;3:219-36.

13. Vealey RS. Current status and prominent issues in sport psychology interventions. Med Sci Sport Exerc. 1994;26:495-502. doi:10.1249/00005768-199404000-00015.

14. Weinberg RS, Comar W. The effectiveness of psychological interventions in competitive sport. Sports Med. 1994;18:406-18. doi:10.2165/00007256-199418060-00005.

15. Martin GL, Thompson K, Regehr K. Studies using single-subject designs in sport psychology: 30 years of research. Behav Analyst. 2004;27:263-80.

16. Martin GL, Vause T, Schwartzman LH. Experimental studies of psychological interventions with athletes in competitions: Why so few? Behav Modif. 2005;29:616-41. doi:10.1177/0145445503259394.

17. Barker JB, Mellalieu SD, McCarthy PJ, et al. A review of single-case research in sport psychology 1997-2012: Research trends and future directions. J Appl Sport Psychol. 2013;25:4-32. doi:10.1080/10413200.2012.709579.

18. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking review in health care. York, UK: York Publishing Services Ltd.; 2009.

19. Egger M, Smith GD, O'Rourke K. Rationale, potentials, and promise of systematic reviews. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: Meta-analysis in context. London, UK: BMJ Books; 2001. p. 3-19.

20. Driskell JE, Copper C, Moran A. Does mental practice enhance performance? J Appl Psychol. 1994;79:481-92. doi:10.1037/0021-9010.79.4.481.

21. Feltz DL, Landers DM. The effects of mental practice on motor skill learning and performance: A meta-analysis. J Sport Psychol. 1983;5:25-57.

22. Hinshaw KE. The effects of mental practice on motor skill performance: Critical evaluation and meta-analysis. Imagin Cogn Pers. 1991;11:3-35. doi:10.2190/X9BA-KJ68-07AN-QMJ8.

23. Kyllo LB, Landers DM. Goal setting in sport and exercise: A research synthesis to resolve the controversy. J Sport Exerc Psychol. 1995;17:117-37.

24. Martin LJ, Carron AV, Burke SM. Team building interventions in sport: A meta-analysis. Sport Exerc Psychol Rev. 2009;5:3-18.

25. Rovio E, Arvinen-Barrow M, Weigand DA, et al. Team building in sport: A narrative review of the program effectiveness, current methods, and theoretical underpinnings. Athletic Insight: The Online Journal of Sport Psychology. 2010;12:147-64.

26. Hatzigeorgiadis A, Zourbanos N, Galanis E, et al. Self-talk and sports performance: A meta-analysis. Perspect Psychol Sci. 2011;6:348-56. doi:10.1177/1745691611413136

27. Tod D, Hardy J, Oliver E. Effects of self-talk: A systematic review. J Sport Exerc Psychol. 2011;33:666-87.

28. Rumbold JL, Fletcher D, Daniels K. A systematic review of stress management interventions with sport performers. Sport Exerc Perform Psychol. 2012;1:173-93. doi:10.1037/a0026628.

29. Dishman RK. Identity crisis in North American sport psychology: Academics in professional issues. J Sport Psychol. 1983;5:123-34.

30. Egger M, Jüni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Asses. 2003;7:1-76.

31. Reeves BC, Deeks JJ, Higgins JPT, et al. Including non-randomized studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. The Cochrane Collaboration; 2011.

32. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. In: Gage NL, editor. Handbook of research on teaching. Boston, MA: Houghton Mifflin Company; 1963. p. 1-76.

33. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: Are implementation effects out of control? Clin Psychol Rev. 1998;18:23-45. doi:10.1016/S0272-7358(97)00043-3.

34. Gresham FM, Gansle KA, Noell GH, et al. Treatment integrity of school-based behavioral intervention studies: 1980-1990. School Psychol Rev. 1993;22:254-72.

35. Waltz J, Addias ME, Koerner K, et al. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. J Consult Clin Psychol. 1993;61:620-30. doi:10.1037/0022-006X.61.4.620.

36. Higgins JPT, Altman DG, Gøtzche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J. 2011;343:d5928. doi:10.1136/bmj.d5928.

37. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration; 2011.

38. Higgins JPT, Deeks JJ, Altman DG. Special topics in statistics. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration; 2011.

39. Borenstein M, Hedges LV, Higgins JPT, et al. Comprehensive meta-analysis: A computer program for research synthesis. 3.0 ed. Englewood, NJ: Biostat; 2014.

40. Bangert-Drowns RL. Review of development in meta-analytic method. Psychol Bull. 1986;99:388-99. doi:10.1037/0033-2909.99.3.388.

41. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV,

editors. The handbook of research synthesis and meta-analysis. 2nd ed. New York City, NY: Russell Sage Foundation; 2009. p. 357-76.

42. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Stat. 1981;6:107-28. doi:10.3102/10769986006002107.

43. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: Meta-analysis in context. London, UK: BMJ Books; 2001. p. 285-312.

44. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.

45. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med. 2003;22:2693-710. doi:10.1002/sim.1482.

46. Egger M, Smith GD. Principles of and procedures for systematic reviews. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: Meta-analysis in context. London, UK: BMJ Books; 2001. p. 23-42.

47. Borenstein M, Hedges LV, Higgins JPT, et al. Introduction to meta-analysis. Chichester, UK: John Wiley & Sons; 2009.

48. Hedges LV, Olkin I. Statistical methods for meta-analysis. New York City, NY: Academic Press; 1985.

49. Borenstein M. Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC, editors. Handbook of research synthesis and meta-analysis. 2nd ed. New York City, NY: Russell Sage Foundation; 2009. p. 221-35.

50. Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. Stat Med. 1988;7:889-94. doi:10.1002/sim.4780070807.

51. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and

adjusting for publication bias in meta-analysis. Biometrics. 2000;56:455-63. doi:10.1111/j.0006-341X.2000.00455.x.

52. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21:1559-73. doi:10.1002/sim.1187.

53. Englert C, Bertrams A, Furley P, et al. Is ego depletion associated with increased distractibility? Results from a basketball free throw task. Psychol Sport Exerc. 2015;18:26-31.

54. Landers DM, Petruzzello SJ, Salazar W, et al. The influence of electrocortical biofeedback on performance in pre-elite archers. Med Sci Sport Exerc. 1991;23:123-9.

55. Reeves JL, Tenenbaum G, Lidor R. Choking in front of the goal: The effects of self-consciousness training. Int J Sport Exerc Psychol. 2007;5:240-54. doi:10.1080/1612197X.2007.9671834.

56. Barker JB, Jones MV, Greenlees I. Assessing the immediate and maintained effects of hypnosis on self-efficacy and soccer wall-volley performance. J Sport Exerc Psychol. 2010;32:243-52.

57. Baudry L, Leroy D, Chollet D. The effect of combined self- and expert-modelling on the performance of the double leg circle on the pommel horse. J Sport Sci. 2006;24:1055-63. doi:10.1080/02640410500432243.

58. Burroughs WA. Visual simulation training of baseball batters. Int J Sport Psychol. 1984;15:117-26.

59. Caserta RJ, Young J, Janelle CM. Old dogs, new tricks: Training the perceptual skills of senior tennis players. J Sport Exerc Psychol. 2007;29:479-97.

60. Donohue B, Miller A, Beisecker M, et al. Effects of brief yoga exercises and motivational preparatory interventions in distance runners: Results of a controlled trial. Br J Sports Med. 2006;40:60-3. doi:10.1136/bjsm.2005.020024.

61. Gabbett TJ, Carius J, Mulvey M. Does improved decision-making ability reduce the physiological demands of game-based activities in field sport athletes? J Strength Cond Res. 2008;22:2027-35.

62. Hazell J, Cotterill ST, Hill DM. An exploration of pre-performance routines, self-efficacy, anxiety and performance in semi-professional soccer. Eur J Sport Sci. 2014;14:603-10.

63. Hill KL, Borden F. The effect of attentional cueing scripts on competitive bowling performance. Int J Sport Psychol. 1995;26:503-12.

64. Kachanathu SJ, Verma SK, Khanna GL. Effect of music therapy on heart rate variability: A reliable marker to pre-competition stress in sports performance. J Med Sci. 2013;13:418-24. doi:10.3923/jms.2013.418.424.

65. Kress J, Schroeder J, Potteiger JA, et al. The use of psychological skills training to increase 10 KM cycling performance: An exploratory investigation. Int Sports J. 1999;3:44-54.

66. Lane A, Streeter B. The effectiveness of goal setting as a strategy to improve basketball shooting performance. Int J Sport Psychol. 2003;34:138-50.

67. Lautenbach F, Laborde S, Mesagno C, et al. Nonautomated pre-performance routine in tennis: An intervention study. J Appl Sport Psychol. 2014;27:123-31. doi:10.1080/10413200.2014.957364.

68. Lorains M, Ball K, MacMahon C. An above real time training intervention for sport decision making. Psychol Sport Exerc. 2013;14:670-4.

69. Madden G, McGown C. The effect of hemisphericity, imagery, and relaxation on volleyball performance. J Hum Movement Stud. 1988;14:197-204.

70. Malouff JM, McGee JA, Halford HT, et al. Effects of pre-competition positive imagery and self-instructions on accuracy of serving in tennis. J Sport Behav. 2008;31:264-75.

71. Mauger AR, Jones AM, Williams CA. Influence of feedback and prior experience on pacing during a 4-km cycle time trial. Med Sci Sport Exerc. 2009;41:451-8. doi:10.1249/MSS.0b013e3181854957.

72. McCann P, Lavallee D, Lavallee R. The effect of pre-shot routines on golf wedge shot performance. Eur J Sport Sci. 2001;1:1-10. doi:10.1080/17461390100071503.

73. Mesagno C, Mullane-Grant T. A comparison of different pre-performance routines as possible choking interventions. J Appl Sport Psychol. 2010;22:343-60. doi:10.1080/10413200.2010.491780.

74. Mesagno C, Hill DM, Larkin P. Examining the accuracy and in-game performance effects between pre- and post-performance routines: A mixed methods study. Psychol Sport Exerc. 2015;19:85-94. doi:10.1016/j.psychsport.2015.03.005.

75. Miller A, Donohue B. The development and controlled evaluation of athletic mental preparation strategies in high school distance runners. J Appl Sport Psychol. 2003;15:321-34. doi:10.1080/10413200390238004.

76. Murgia M, Sors F, Muroni AF, et al. Using perceptual home-training to improve anticipation skills of soccer goalkeepers. Psychol Sport Exerc. 2014;15:642-8. doi:10.1016/j.psychsport.2014.07.009.

77. Olsson CJ, Jonsson B, Nyberg L. Internal imagery training in active high jumpers. Scand J Psychol. 2008;49:133-40. doi:10.1111/j.1467-9450.2008.00625.x.

78. Oudejans RRD, Heubers S, Ruitenbeek J-RJAC, et al. Training visual control in wheelchair basketball shooting. Res Q Exerc Sport. 2012;83:464-9. doi:10.1080/02701367.2012.10599881.

79. Panchuk D, Farrow D, Meyer T. How can novel task constraints be used to induce acute changes in gaze behaviour? J Sport Sci. 2014;32:1196-201. doi:10.1080/02640414.2013.876089.

80. Paul M, Garg K. The effect of heart rate variability biofeedback on performance psychology of basketball players. Appl Psychophysiol Biofeedback. 2012;37:131-44. doi:10.1007/s10484-012-9185-2

81. Paul M, Garg K, Sandhu JS. Role of biofeedback in optimizing psychomotor performance in sports. Asian J Sports Med. 2012;3:29-40.

82. Predebon J, Docker SB. Free-throw shooting performance as a function of preshot routines. Percept Motor Skill. 1992;75:167-71.

83. Ramsey R, Cumming J, Edwards MG, et al. Examining the emotion aspect of PETTLEP-based imagery with penalty taking in soccer. J Sport Behav. 2010;33:295-314.

84. Shafizadeh M, Platt GK. Effect of verbal cueing on trajectory anticipation in the penalty kick among novice football goalkeepers. Percept Motor Skill. 2012;114:174-84.

85. Shelton TO, Mahoney MJ. The content and effect of "psyching-up" strategies in weight lifters. Cognitive Ther Res. 1978;2:275-84.

86. Solberg EE, Berglund K-A, Engen Ø, et al. The effect of meditation on shooting performance. Br J Sports Med. 1996;30:342-6.

87. Weinberg RS, Seabourne TG, Jackson A. Effects of visuo-motor behavior rehearsal, relaxation, and imagery on karate performance. J Sport Psychol. 1981;3:228-38.

88. Wrisberg CA, Anshel MH. The use of positively-worded performance reminders to reduce warm-up decrement in the field hockey penalty shot. J Appl Sport Psychol. 1997;9:229-40. doi:10.1080/10413209708406484.

89. Bertakis KD. The influence of gender on the doctor-patient interaction. Patient Educ Couns. 2009;76:356-60. doi:10.1016/j.pec.2009.07.022.

90. Sandhu H, Adams A, Singleton L, et al. The impact of gender dyads on doctor-patient communication: A systematic review. Patient Educ Couns. 2009;76:348-55. doi:10.1016/j.pec.2009.07.010.

91. Leach MJ. Rapport: A key to treatment success. Complement Ther Clin Pract. 2005;11:262-5. doi:10.1016/j.ctcp.2005.05.005.

92. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. Br Med J. 2009;339:b2700. doi:10.1136/bmj.b2700.

93. Bushman BJ, Wang MC. A procedure for combining sample standardized mean differences and vote counts to estimate the population standardized mean difference in fixed event models. Psychol Methods. 1996;1:66-80. doi:10.1037/1082-989X.1.1.66.

94. Bushman BJ, Wang MC. Vote counting methods in meta-analysis. In: Cooper HM, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis. 2nd ed. New York City, NY: Russell Sage Foundation; 2009. p. 207-20.

95. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J. 2010;340. doi:10.1136/bmj.c869.

96. American Psychological Association. Publication manual of the American Psychological Association. 6th ed. Washington, DC: American Psychological Association; 2010.

97. Ardito RB, Rabellino D. Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. Front Psychol. 2011;2:e270. doi:10.3389/fpsyg.2011.00270.

98. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Br Med J. 2009;339:b2535. doi:10.1136/bmj.b2535.

Table 1

Descriptive analysis of study characteristics

Characteristic	Frequency (*k*)
Total number of studies	35
Study design	
RCT	34
RXCT	1
Matching	
Unmatched	29
Matched	6
Risk of bias	
Random sequence generation	
No (high risk)	3
Unclear	0
Yes (low risk)	32
Allocation concealment	
No (high risk)	0
Unclear	33
Yes (low risk)	2
Blinding of participants and personnel	
No (high risk)	33
Unclear	2
Yes (low risk)	0
Blinding of outcome assessment	
No (high risk)	3
Unclear	1
Yes (low risk)	31
Incomplete outcome data adequately addressed	
No (high risk)	3
Unclear	0
Yes (low risk)	32
Free from suggestion of selective outcome reporting	
No (high risk)	3
Unclear	0
Yes (low risk)	32
Participant characteristics	
Sex	
Female only	3
Male only	17
Mixed	12
Not reported	3
Competitive standard	
Local	15
Regional	2
National	6
International	1

Mixed	11
Type of sport	
Archery	1
Australian Football	2
Baseball	1
Basketball & wheelchair basketball	5
Cycling	2
Field hockey	1
Golf	2
Gymnastics	1
Martial arts (karate)	1
Shooting	2
Soccer	7
Tennis	3
Ten-pin bowling	2
Track and field	3
Volleyball	1
Weightlifting	1
Intervention characteristics	
Intervention type[a]	
Activation (P)	2
Attentional focus manipulation (P)	5
External statements (P)	1
Feedback (P, PS)	5
Goal setting (P)	3
Hypnosis (P)	1
Imagery (P)	5
Motivational video footage (PS)	2
Multi-intervention (P, PS)	4
Multimodal pre-performance routine (P)	8
Perceptual training (P, PS)	9
Post-performance questioning	1
Relaxation (P)	5
Self-talk (P, PS)	6
Stress inoculation (P)	1
Intervention components	
Single	35
Multiple	23
Intervention provider	
Self	6
Coach	6
Researcher	36
Practitioner	4
Equipment	6
Intervention setting	
Individual	19
Group	39
Performance outcome characteristic	
Component of fitness	
Muscular strength	1

Speed and power	3
Overall/competition outcome	
Bowling performance	4
Endurance performance	4
High jump performance	4
Karate fight performance	3
On-pitch decision making	3
Target shooting	4
Tennis service return	3
Volleyball in-game service accuracy	2
Technical task[b]	
Australian Football set shot drill	1
Basketball drill	8
Off-pitch decision making (e.g., goalkeeper shot location prediction)	7
Field hockey penalty stroke drill	6
Golf shot drill	4
Gymnastics skill	4
Soccer drill	7
Tennis serve drill	3
Volleyball pass drill	1

Note. RCT = parallel individual randomized controlled trial; RXCT = cross-over individual randomized controlled trial; P = psychological intervention; PS = psychosocial intervention;

[a]No suitable social interventions were identified in the literature search. Some intervention types included variations that met both psychological and psychosocial classifications.

[b]Technical tasks performed under low and high pressure were considered distinct tasks and are included individually.

Table 2

Study effect size statistics

		Effect size		
Analysis	*n*	Hedges' *g*	SE	95% CI
Posttest				
Barker et al. [56]	59	0.48	0.26	[-0.03, 0.99]
Baudry et al. [57]	16	1.08	0.51	[0.08, 2.08]
Burroughs [58] Experiment 1	23	0.95	0.43	[0.12, 1.79]
Caserta et al. [59]	18	6.94	1.28	[4.43, 9.45]
Donohue et al. [60]	60	0.07	0.26	[-0.43, 0.57]
Gabbett et al. [61]	16	0.26	0.48	[-0.67, 1.20]
Hazell et al. [62]	20	1.00	0.46	[0.11, 1.90]
Hill & Borden [63]	31	0.51	0.36	[-0.19, 1.21]
Kachanathu et al. [64]	99	0.66	0.21	[0.26, 1.06]
Kress et al. [65]	11	0.47	0.56	[-0.63, 1.58]
Landers et al. [54]	16	0.02	0.47	[-0.90, 0.95]
Lane & Streeter [66]	36	0.20	0.33	[-0.44, 0.84]
Lautenbach et al. [67]	29	-0.53	0.37	[-1.26, 0.19]
Lorains et al. [68]	30	6.00	0.86	[4.32, 7.69]
Madden & McGown [69]	17	0.15	0.46	[-0.75, 1.06]
Malouff et al. [70]	76	0.57	0.23	[0.11, 1.02]
Mauger et al. [71]	18	1.13	0.49	[0.18, 2.08]
McCann et al. [72]	18	0.90	0.48	[-0.05, 1.85]
Mesagno et al. [74]	17	0.23	0.47	[-0.69, 1.15]
Mesagno & Mullane-Grant [73]	24	1.13	0.43	[0.29, 1.97]
Miller & Donohue [75]	60	0.07	0.26	[-0.43, 0.57]
Murgia et al. [76]	25	1.02	0.41	[0.21, 1.83]
Olsson et al. [77]	19	0.13	0.44	[-0.74, 0.99]
Oudejans et al. [78]	12	0.96	0.58	[-0.17, 2.08]
Panchuk et al. [79]	14	-0.97	0.56	[-2.07, 0.14]
Paul & Garg [80]	20	1.08	0.48	[0.13, 2.03]
Paul et al. [81]	20	0.74	0.46	[-0.16, 1.63]
Predebon & Docker [82]	20	0.31	0.43	[-0.54, 1.16]
Ramsey et al. [83]	22	1.35	0.46	[0.45, 2.26]
Reeves et al. [55]	25	-0.19	0.40	[-0.97, 0.59]
Shafizadeh & Platt [84]	28	0.31	0.37	[-0.42, 1.03]
Shelton & Mahoney [85]	28	0.39	0.37	[-0.34, 1.12]
Solberg et al. [86]	25	-0.04	0.41	[-0.83, 0.76]
Weinberg et al. [87]	16	0.46	0.48	[-0.48, 1.41]
Wrisberg & Anshel [88]	30	0.06	0.36	[-0.64, 0.77]
First Follow-up[a]				

Barker et al. [56]	28	0.67	0.38	[-0.07, 1.41]
Lane & Streeter [66]	36	0.22	0.33	[-0.43, 0.87]
Lorains et al. [68]	30	3.72	0.61	[2.51, 4.92]
McCann et al. [72]	18	0.48	0.47	[-0.43, 1.39]
Mesagno et al. [74]	17	0.47	0.47	[-0.45, 1.39]
Paul & Garg [80]	20	1.74	0.58	[0.61, 2.87]
Paul et al. [81]	20	1.58	0.55	[0.51, 2.65]
Predebon & Docker [82]	20	1.07	0.46	[0.16, 1.97]
Second Follow-up[b]				
Lane & Streeter [66]	36	-0.13	0.33	[-0.77, 0.51]
Predebon & Docker [82]	20	0.50	0.44	[-0.36, 1.35]

Note. *SE* = standard error; CI = confidence interval.

[a]Follow-up assessments ranged from two to six weeks after pre-test (one to four weeks after the intervention finished). [b]Follow-up assessments ranged from three to six weeks after pre-test (two to four weeks after the intervention finished).

Table 3

Model summary statistics

		Mean summary effect[a]			Heterogeneity statistics								
Model	*k*	Hedges' *g*	95% CI	95% PI	*Q*	*p*-value	T^2	95% CI_{T^2}	*T*	95% CI_T	I^2	95% CI_{I^2}
1	35	0.57	[0.22, 0.92]	[-0.68, 1.82]	113.00	< 0.001	0.35	[0.20, 0.55]	0.59	[0.45, 0.74]	69.91	[57.70, 78.60]
2	33	0.43	[0.26, 0.59]	[-0.10, 0.95]	45.48	0.058	0.06	[0.00, 0.16]	0.24	[0.00, 0.42]	29.64	[0.00, 53.50]
3	32	0.62	[0.23, 1.00]	[-0.71, 1.94]	110.69	< 0.001	0.39	[0.23, 0.61]	0.62	[0.48, 0.78]	71.99	[60.27, 80.26]
4	8	1.02	[-0.37, 2.41]	[-1.72, 3.77]	43.59	< 0.001	0.91	[0.42, 1.83]	0.96	[0.64, 1.35]	83.94	[70.44, 91.28]
5	8	1.16	[0.25, 2.08]	[-1.12, 3.45]	31.61	< 0.001	0.72	[0.27, 1.59]	0.85	[0.52, 1.26]	77.45	[57.21, 88.54]
6	7	0.78	[0.27, 1.30]	[-0.19, 1.76]	9.02	0.173	0.10	[0.00, 0.42]	0.31	[0.00, 0.65]	33.45	[0.00, 68.13]
7	7	1.32	[0.30, 2.35]	[-1.15, 3.80]	24.96	< 0.001	0.75	[0.25, 1.79]	0.87	[0.50, 1.34]	75.96	[50.82, 88.25]

Note. Lower limits for 95% CI $_{T^2}$, 95% CI $_T$, and 95% CI $_{I^2}$ set at zero where returned values were negative. Model 1 = posttest comparisons, all studies included; Model 2 = posttest comparisons, potential outlier studies [59, 68] excluded; Model 3 = posttest comparisons, high-risk to reporting bias studies [61, 66, 88] excluded; Model 4 = posttest comparisons, only studies with follow-up comparisons included; Model 5 = follow-up comparisons, all studies included; Model 6 = follow-up comparisons, potential outlier studies [68] excluded; Model 7 = follow-up comparisons, high-risk to reporting bias studies [66] excluded. *k* = number of studies, CI = confidence interval; PI = predictive interval; *Q* = total variance in model; *T* = estimate of tau based on observed effects; I^2 = proportion of the observed variance that is explained by the between-study differences.

[a]Computed using a random effects (method of moments) model with Knapp-Hartung [45] modification.

Table 4

Meta-regression model summaries

Model	T^2	R^2	Test of model statistics				Goodness of fit statistics		
			F	*df*1	*df*2	*p*-value	*Q*	*df*	*p*-value
1: Full posttest model[a]	0.35	0.00					113.00	34	< 0.001
Sex									
8: Intercept only[b]	0.34	0.00					104.77	31	< 0.001
9: Covariate[b, c]	0.33	0.04	0.86	1	30	0.360	98.37	30	< 0.001
Competitive standard									
10: Intercept only[d]	0.23	0.00					63.52	23	< 0.001
11: Covariate[d, e]	0.26	0.00	0.33	3	20	0.802	60.73	20	< 0.001
Intervention provider									
12: Intercept only[f]	0.36	0.00					110.44	33	<0.001
13: Covariate[f,g]	0.34	0.05	1.45	4	29	0.244	91.99	29	<0.001
Intervention components									
14: Intercept only[h]	0.41	0.00					106.00	28	<0.001
15: Covariate[h,i]	0.42	0.00	1.41	1	27	0.245	102.66	27	<0.001
Type of intervention									
16: Intercept only[j]	0.39	0.00					108.34	30	<0.001
17: Covariate[j,k]	0.32	0.20	9.08	1	29	0.005	90.62	29	<0.001
Outcome characteristics									
18: Intercept only[l]	0.36	0.00					112.52	33	<0.001
19: Covariate[l,m]	0.38	0.00	0.15	2	31	0.862	110.69	31	<0.001
Combined (sex, intervention provider, type of intervention)									
20: Intercept only[n]	0.38	0.00					100.33	28	<0.001
21: Covariate[n,o]	0.27	0.28	2.96	6	22	0.028	62.49	22	<0.001

Note. Random effects (method of moments) computational model with Knapp-Hartung [45] modification, Hedges' *g* effect size estimate. T^2 = measure of variance of true effects; R^2 = proportion of total between-study variance explained by the model; *F* = F-ratio; *Q* = total variance in model.
[a]*k* = 35, *n* = 997. [b]*k* = 32, *n* = 950, studies not reporting sex were excluded. [c]Covariate = % male participants in sample. [d]*k* = 24, *n* = 783, studies containing mixed or not reported competitive standard were excluded. [e]Covariate = competitive standard (local, regional, national, international) with local level set as the comparison group. [f]*k* = 34, *n* = 937, studies including different types of provider were excluded. [g]Covariate = intervention provider, with self-delivered interventions set as the comparison group. [h]*k* = 29, *n* = 840, studies involving both single and multiple component interventions were excluded. [i]Covariates = single or multiple component intervention, with single

component set as the comparison group. [j]$k = 31$, $n = 880$, studies including different types of interventions were excluded. [k]Covariates = type of intervention, with psychological interventions set as the comparison group. [l]$k = 34$, $n = 980$, studies including multiple types of outcome characteristic were excluded. [m]Covariates = type of outcome characteristic, with component of fitness outcomes set as the comparison group. [n]$k = 29$, $n = 850$, studies not reporting sex, including multiple providers and/or including multiple types of intervention were excluded. [o]Covariates = % male participants in sample, intervention provider, and type of intervention, with self and psychological set as the comparison groups.

Fig. 1 Study flow diagram following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [98]. *n* = number of papers; *k* = number of individual studies.

Fig. 2 Galbraith diagram of randomized controlled trial studies assessed at posttest. Each study effect estimate divided by its standard error (SE) is plotted against the reciprocal of its standard error (weight).

Fig. 3 Funnel plot of randomized controlled trial studies at posttest displaying observed data points and adjusted values based on Duval and Tweedie's trim and fill procedure. Observed data are shown by the clear shapes and the imputed data are shown by the filled shapes; no imputed data were added in this analysis. The clear diamond displays the mean summary effect for observed data and the filled diamond displays the estimated mean summary effect after imputed data added. Inverse of standard error (SE) for each study was plotted against Hedges' *g* using values from a random effects model.

Fig. 4 Galbraith diagram of randomized controlled trial studies assessed at follow-up. Each study effect estimate divided by its standard error (SE) is plotted against the reciprocal of its standard error (weight).

Fig. 5 Funnel plot of randomized controlled trial studies at follow-up displaying observed data points and adjusted values based on Duval and Tweedie's trim and fill procedure. Observed data are shown by the clear shapes and the imputed data are shown by the filled shapes; no imputed data were added in this analysis. The clear diamond displays the mean summary effect for observed data and the filled diamond displays the estimated mean summary effect after imputed data added. Inverse of standard error (SE) for each study was plotted against Hedges' *g* using values from a random effects model.

Fig. 6 Plot of Hedges' *g* for participant characteristics: (a) sex (% of males in sample); (b) competitive standard. Lines indicate predicted effect size and 95% confidence interval using a random effects model, and circles represent studies with circle size proportionate to the study weight.

Fig. 7 Plot of Hedges' *g* for interventions characteristics: (a) main provider of intervention; (b) intervention components; (c) type of intervention. Lines indicate predicted effect size and 95% confidence interval using a random effects model, and circles represent studies with circle size proportionate to the study weight.

Fig. 8 Plot of Hedges' *g* for outcome characteristic (type of performance outcome). Lines indicate predicted effect size and 95% confidence interval using a random effects model, and circles represent studies with circle size proportionate to the study weight.