The Effectiveness of Psychological Skills Training and Behavioral Interventions in Sport using Single-Case Designs: A Meta Regression Analysis of the Peer-Reviewed Studies

Jamie B. Barker[1], Matthew J. Slater[2], Geoff Pugh[3], Stephen D. Mellalieu[4], Paul J. McCarthy[5], Marc V. Jones[6], & Aidan Moran[7]

Date of Submission:

July 20th 2018

Accepted:

11th June 2020

Author Note: [1]School of Sport, Exercise and Health Sciences, Loughborough University, United Kingdom; [2]School of Life Sciences and Education, Staffordshire University, United Kingdom; [3]Centre for Applied Business Research, Staffordshire University; [4]Cardiff School of Sport and Health Sciences Cardiff Metropolitan University, Wales; [5] Department of Psychology and Allied Health Sciences, Glasgow Caledonian University; [6] Department of Psychology, Manchester Metropolitan University [7] School of Psychology, University College Dublin, Ireland.

Correspondence concerning this article should be addressed to Jamie B. Barker, School of Sport, Exercise and Health Sciences, Loughborough University, UK, LE11 3TU. Tel: +44 1509 226 302. Electronic mail may be sent to j.b.barker@lboro.ac.uk

Abstract

We used a novel meta regression analysis approach to examine the effectiveness of psychological skills training and behavioral interventions in sport assessed using single-case experimental designs (SCEDs). One hundred and twenty-one papers met the inclusion criteria applied to eight database searches and key sport psychology journals. Seventy-one studies reported sufficient detail for effect sizes to be calculated for the effects of psychological skills training on psychological, behavioral, and performance variables. The unconditional mean effect size for weighted ($\Delta = 2.40$) and unweighted ($\Delta = 2.83$) models suggested large improvements in psychological, behavioral, and performance outcomes associated with implementing cognitive-behavioral psychological skills training and behavioral interventions with a SCED. However, meta-regression analysis revealed important heterogeneities and sources of bias within this literature. First, studies using a group-based approach reported lower effect sizes compared to studies using single-case approaches. Second, the single-case studies, (over 90 per cent the effect sizes), revealed upwardly biased effect sizes arising from: (i) positive publication bias such that studies using lower numbers of baseline observations reported larger effects, while studies using larger numbers of baseline observations reported smaller – but still substantial – effects; (ii) not adopting a multiple baseline design; and (iii) not establishing procedural reliability. We recommend that future researchers using SCED's should consider these methodological issues.

Keywords: meta regression analysis, psychological skills training, single-case experimental designs, procedural reliability, applied sport psychology

The Effectiveness of Psychological Skills Training and Behavioral Interventions in Sport Using Single-Case Designs: A Meta Regression Analysis of the Peer-Reviewed Studies

The growth, development, and professionalism of sport psychology have further increased the necessity for those working in applied settings to demonstrate accountability and the need for evidence-based practice (Anderson, Miles, Mahoney, & Robinson, 2002; Gardner & Moore, 2006; Hanton & Mellalieu, 2012). Specifically, accountability through evidence-based practice and intervention evaluation and effectiveness is one of the most pressing and essential professional practice issues underpinning further growth of our discipline (e.g., Barker, Mellalieu, McCarthy, Jones, & Moran, 2013; Gardner & Moore, 2006). Interventions in applied sport psychology typically occur through the model of psychological skills training, referring to the "systematic and consistent practice of mental or psychological skills for the purpose of enhancing performance, increasing enjoyment, or achieving greater sport and physical activity satisfaction" (p. 230; Weinberg, 2019). Further, Vealey (1994) emphasized the importance of athletes developing cognitive skills to manage the demands of sport. Although behavioral interventions have largely the same purpose as psychological skills training, they differ in nature by focussing on techniques to modify, alter, or redirect behavior (e.g., public posting of athlete attendance; Michie et al., 2013). The ability to demonstrate objective performance improvements through behavioral change as a direct consequence of psychological skills training and behavioral interventions is an essential facet of sport psychology research that has, in the past, not always been effectively demonstrated (cf. Hardy & Jones, 1994; Smith, 1989).

The important role that cognition plays in psychological skills training, building on earlier behavioral interventions, gave rise to the cognitive-behavioral approach to behavior modification which came to the fore in the 1970's (e.g., Mahoney, 1974; Meichenbaum, 1977). Applications of the cognitive behavioral approach included Visuo-motor Behaviour Rehearsal (Suinn, 1972), Cognitive-affective Stress Management Training programme (Smith, 1980), and Stress Inoculation Training (Meichenbaum, 1977). See Mace (1990) for a review of these intervention programmes. While there are different types of techniques underpinned by cognitive-behavioral principles (e.g., Rational Emotive Behaviour Therapy, REBT; Ellis, 1957, Cognitive-Behavior Modification; Meichenbaum, 2010), they share the central premise that cognitive mediators influence psychological and behavioral responses (Wessler, 1986). Based on this approach, the role of cognition is central in determining an athlete's response to situations because it is how they perceive the demands of the environment (Mahoney, 1974), and appraise their ability to cope (Lazarus, Coyne, & Folkman, 1984), that determines their psychology and behavior, ultimately guiding their performance.

Determining causality in applied sport psychology has often been fraught with problems. These issues include the use of research designs that lack internal or external validity (or both), a failure to assess practical or clinical as opposed to statistical significance, and the use of performance measures that have been too global in nature (Hrycaiko & Martin, 1996; Martin et al., 2005). Attempts to alleviate such concerns have typically been in the form of review or meta-analysis studies, that have generally revealed some positive effects of psychological skills training, but these effects are dependent on factors such as study design and type of psychological skills training. For example, Greenspan and Feltz (1989) provided an overall examination of the effectiveness of psychological skills training used with athletes. In general, the interventions underpinned by cognitive behavioral principles (e.g., cognitive restructuring) used to enhance athletes' performance in competitive situations were associated with some improvements, yet positive effects were seen in less than half the 23 studies. Furthermore, Martin et al. (2005) noted that with so few published experimental studies, generalizations could only be offered with caution. Although 14 out of the 15 studies included interventions which had a positive effect, only 9 highlighted substantial intervention effects with no studies measuring follow-up intervention effects. Reviews documenting the effects of specific psychological skills training (e.g., goal setting) in relation to sport performance and psychological outcomes have yielded similar positive results (see Burton, Naylor, & Holliday, 2001; Kyllo & Landers, 1995; Rumbold, Fletcher, & Daniels, 2012). For example, Tod, Hardy, and Oliver (2011) completed a systematic review examining the relationship between self-talk and performance in 47 studies and supported the beneficial use of self-talk strategies on performance (e.g., positive self-talk improved performance). More recently, there has been meta-analytical support for the positive and moderate effects of psychological and psycho-social interventions (e.g., pre-performance routines and perceptual training) on sport performance (Brown & Fletcher, 2017).

Collectively, these data indicate partial support for the effectiveness of a myriad of psychological skills training techniques (including relaxation, imagery, goal-setting, arousal regulation, self-talk, and stress management) used in real-world sport settings. While these reviews have highlighted the broad range of psychological skills training interventions, there are additional techniques, including hypnosis (Barker & Jones, 2006; 2008) and REBT (Turner & Barker, 2013; Turner & Davis, 2019), that have gained attention from sport psychologists. Aligned with the current definition of psychological skills training, hypnosis and REBT use cognitive and affective strategies to bring about changes in psychological, behavioral, and performance outcomes. However, taken together, these data do not conclusively demonstrate the efficacy and effectiveness of psychological skills training (Smith, 1989; Vealey, 1994). Possible reasons for such equivocal findings are related to the types of methods, including the research design, used to determine intervention effectiveness (Martin et al., 2005; Smith, 1989; Vealey, 1994). Typically, intervention studies have sought to determine effectiveness through “nomothetic” (i.e., concerning the formulation of general laws) methodology involving experimental designs and multivariate analyses (e.g., Martin et al. 2005). Such methodology, while minimizing threats to internal validity makes it difficult to glean “idiographic” (i.e., pertaining to individual cases) intervention responses and patterns (Kazdin, 1982). Although nomothetic designs have an important theoretical and psychometric development function, they do not allow for the detailed and objective exploration of individuals in real-world settings – a fact which hinders understanding of intervention efficacy and effectiveness (e.g., Barker, McCarthy, Jones, & Moran, 2011; Kazdin, 2011; Meredith, Dicks, Noel, & Wagstaff, 2018; Smith, 2012). Accordingly, single-case experimental designs (SCEDs) offer a viable means of maintaining scientific rigor in applied settings while providing a platform for examining the idiographic processes and outcomes of psychological and behavioral intervention effects across time with individuals and groups (e.g., Barlow, Nock, & Hersen, 2009; Meredith et al., 2018; Morgan & Morgan, 2009).

A unique feature of SCEDs is the capacity to conduct experimental investigations with one or a few cases and the ability to rigorously evaluate individual nuances and effects of interventions between baseline and post intervention phases (Kazdin, 2011). SCEDs are not considered replacements for more traditional controlled group designs but are a complementary and/or an alternative approach when developing new intervention protocols or working with small or unique populations. SCEDs enable the detection of intervention effects for individuals who would otherwise have their nuances masked in a non-significant group design (Barker et al., 2013). A key indicator for determining study quality in SCED’s is that of procedural reliability. Researchers adopting procedural reliability ensure that an intervention is applied and delivered as intended and consistently across participants. Accordingly, SCEDs with procedural reliability can be considered of a better quality than those without (Kazdin, 2011).

While SCEDs do provide a platform for exploring intervention effects, they have certain weaknesses. First, they are insensitive to interaction effects between participants at a study level. Second, given the challenges of statistical analyses it is difficult to determine any quantitative index of confidence in the generalizability of the results. Third, it can be difficult to interpret intervention effects if the baseline shows excessive variability. For this reason, researchers need to establish stable and lengthy baselines of dependent variables before interventions are applied. Finally, although SCEDs are helpful in exploring effects at an individual level, their capacity to generalize findings validly to other participants and settings is questionable (Barker et al., 2011).

The use of SCEDs is supported by substantial evidence that has accepted and adopted SCEDs extensively in behavioral medicine and in clinical settings, health, education, schools, rehabilitation, counseling psychology, and sport (see Smith, 2012). During the past 30 years, sport psychology researchers have repeatedly been encouraged to use and publish SCEDs in relevant journals (e.g., *Case Studies in Sport and Exercise Psychology*) to further advance knowledge of intervention effectiveness and evidence-based practice (e.g., Barker et al., 2013; Bryan, 1987; Hrycaiko & Martin, 1996; Martin, Thomson, & Regehr, 2004). Despite this demand, relatively few SCEDs have been published in sport psychology (see Meredith et al., 2018). Based on a review of 66 studies between 1997-2012, Barker and colleagues (2013) proposed important considerations for SCED researchers. First, there was a sampling reliance on collegiate and recreational athletes, with few studies using professional and/or elite (both able-bodied and disabled) athletes. Second, the multiple-baseline across-participants design was the most frequently used single-case variation, which reflects good practice within SCED research (Kazdin, 2011); however, few designs assessed follow-up or maintenance effects. Third, various psychological (e.g., anxiety, self-confidence) and behavioral (e.g., inappropriate on-court outbursts) outcomes were assessed across the studies, while only 42 of the sampled studies provided detail regarding the key principle of procedural reliability (i.e., the extent to which components of an intervention are consistently delivered across individuals or settings). In addition, it was not evident in the review to what extent the psychological skills training which used SCEDs were effective (i.e., bringing about meaningful changes in target variables). Therefore, adopting a meta-analytic approach to glean such insight would make a significant contribution to the extant literature.

Meta-analysis was designed to yield valid estimates of representative effects from empirical literatures that report large numbers of quantitative results. Yet, in empirical literatures in the life and social sciences, the fog of heterogeneous results often makes it difficult to discern representative effects (Stanley et al., 2013). Accordingly, meta-regression analyses use regression analysis of the primary literature to identify potential sources of variation in research findings, which typically arise from differences in the context and samples of studies or in the design of studies (Stanley & Doucouliagos, 2012; Stanley et al., 2013). The benefits of this statistical approach are two-fold: (i) enabling sources of heterogeneity to be controlled when estimating the representative effect size from a literature; while (ii) simultaneously yielding more fine-grained information on the effects associated with different types of sample (e.g., by sport or standard) or different research designs, procedures, and/or interventions (e.g., multiple baseline and procedural reliability). Meta regression analysis also accounts for publication bias, which is an endemic threat to the validity of quantitative findings in the life and social sciences. As such, larger and more significant effects are over-represented, so that, in a typical quantitative literature: “publication selection biases a literature’s average reported empirical effect away from zero” (Stanley, 2008, p. 104). For the bio-medical sciences, Ioannidis (2005) contended that quantitative research findings in many scientific fields may often be a measure of the prevailing bias, where bias is considered to be the combination of various factors (e.g., exercising discretion over design and/or analysis) that typically leads to conclusions that are not, in fact, defensible, or ‘real’, in the sense of Type I errors (i.e., rejecting the null hypothesis when it is actually true). Furthermore, bias should not be confused with chance variability that causes some findings to be false by chance even though all elements of the study are robust. In contrast, selective or distorted reporting (e.g., of data or analyses) are typical forms of such bias. Indeed, researchers have concluded that publication bias is pervasive across the field of psychology (Kühberger, Fritz, & Scherndl, 2014). The consequences of publication bias are not visible at the level of the individual primary study, yet leave their trace in the literature as a whole. Accordingly, a major contribution of meta regression is to identify the extent to which publication bias exists in the literature; and, simultaneously, to control for publication bias so that a representative effect size can be estimated net of – or “beyond” – publication bias (Stanley, 2005; 2008; Stanley & Doucouliagos, 2012).

In the context of sport psychology interventions adopting SCEDs, meta regression may contribute to our understanding of the peer reviewed literature on at least three levels. First, it facilitates identification of the degree to which publication bias is evident in SCED literature. Second, it reveals the extent to which the heterogeneous reported effects sizes can be explained by the heterogeneity of samples and research designs (i.e., such as athlete standard, research design, or individual vs. multiple mental skill) used. Third, it provides insight into the meaningfulness of change – by identifying and controlling for publication bias and heterogeneous effects in the primary literature, thereby better estimating the representative effect size for SCEDs in applied sport psychology. Exploiting these strengths, the purpose of our current study was to extend the review by Barker et al. (2013) by exploring the overall effectiveness of psychological skills training and behavioral interventions – underpinned by cognitive behavioral principles – using SCEDs through meta-regression analysis. We aimed to answer our research question: "Are psychological skills training programmes and behavioral interventions assessed using SCEDs effective in sport?" Support for this intervention approach in our meta-regression would provide robust evidence, while findings to the contrary would potentially undermine the application of psychological skills training and behavioral interventions using SCEDs in sport.

Method

Inclusion Criteria

Studies that met the following criteria were included: (1) used a single-case methodology – as our research question focussed on interventions that have adopted SCEDs only; (2) published in the English language; (3) peer-reviewed journal publication – as a marker of research quality; (4) a study that applied psychological skills training and/or a behavioral intervention in sport – as our research question focussed on effectiveness of psychological skills training (Weinberg, 2019) and behavioral interventions in sport only; and (5) was a quantitative study of intervention effects – as is the purpose of SCEDs along with the requirement for numerical data for meta-regression analysis.

Search Strategy

In-line with the PRISMA checklist (see supplementary file) we undertook the following procedures. To identify studies that met the inclusion criteria, five databases were searched: PsychARTICLES; PsychINFO; Science Direct; SCOPUS; and SportDiscus. Further, key journals within the SCED literature were searched (e.g., *Journal of Applied Behavioral Analysis* and *Journal of Applied Sport Psychology*). The individual search terms were developed by the authors, and the following were used to identify studies: "*single-case AND sport*"; and "*sport psychology intervention*". In the first instance, the titles were screened and then the abstract of any papers that met the criteria was read. Next, the full manuscript was read to determine whether or not the paper met the criteria. The first and second authors completed the search strategy before cross-referencing with the third author. For example, in SCOPUS the search "*single-case AND sport*" returned 179 titles, and "*sport psychology intervention*" returned 1,400 titles. The search was on-going until January 2019. Finally, the compiled table of studies was shared with all authors for verification and comments. In total, 121 papers met the inclusion criteria. The study selection process can be seen in the PRISMA flow diagram in Figure 1 (cf. Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

Effect Size Calculations

Glass's delta includes the baseline, rather than the pooled, standard deviation and therefore was chosen as the appropriate effect size, because in SCEDs participants act as their own control (Barker et al., 2011). Of the 121 manuscripts, twelve studies reported effect sizes, nearly half of the studies reported or displayed (in graphical form) the necessary values to calculate the effect size (i.e., means and standard deviations for baseline and intervention phases; $n = 59$), while the remaining studies did not report sufficient detail for effect sizes to be calculated ($n = 50$). To achieve a standardised figure, we calculated Glass's delta by hand across the 71 eligible studies. Given that the purpose of psychological interventions may be to increase (e.g., self-efficacy) or decrease (e.g., number of on-court outbursts) variables, the effect sizes were transformed to ensure that positive values represented improvements and negative values detrimental effects.

Effect sizes were calculated for psychological, behavioral, and performance variables across the 71 studies (a total of 367 athletes) resulting in 962 effect sizes (Table 1 shows the study characteristics of the 71 articles). For each study, effect sizes were weighted to eliminate bias towards studies reporting a greater number of effect sizes (e.g., administering multiple questionnaires to athletes). Accordingly, for each study, effect sizes were weighted by the inverse of the number of effect sizes, so that for each study the effect size weights sum to one (e.g., Freeman, Rees, & Hardy, 2009 reported 15 effect sizes and thus was weighted at 1/15 = .067). Both weighted and unweighted models are reported. The observations were filtered, first, to those that related to the A-B phase within SCEDs (n = 648) and, second, to targeted dependent variables (n = 626) rather than control variables. Thus, 626 effect sizes were used in the meta-analysis.

Psychological Skills Training Techniques

A broad range of psychological skills training techniques were used across the 71 studies. The most prevalent were: imagery (n = 15 as an individual mental skill, n = 9 as part of a multiple mental skills package), goal setting (n = 4 as an individual mental skill, n = 8 as part of a multiple mental skills package), self-talk (n = 3 as an individual mental skill, n = 6 as part of multiple skills package), hypnosis (n = 6 as an individual mental skill, n = 2 as part of multiple skills package), and REBT (n = 7 as a multiple skills package, n = 1 with the addition of Personal-Disclosure Mutual-Sharing).

Preliminary Meta-Analysis Procedure

The “funnel plot” of estimated effect sizes (horizontal axis) against the precision of each estimate (vertical axis) is one of the most widely used graphical tools for summarising and describing quantitative literatures and is particularly useful for revealing publication bias (Stanley & Doucouliagos, 2012). A literature without publication bias will yield a symmetrical scatter of observations resembling an inverted funnel; in this case, the mouth of the funnel shows a wide and random scatter of low-precision estimates around the true or authentic effect size; and, as precision increases, the scatter narrows to a spout of high-precision estimates increasingly close to the true effect. Conversely, asymmetry towards the mouth or base indicates publication bias in the literature: in particular, whereas low-precision estimates should be distributed randomly around the true effect size, relatively underpopulated or relatively overpopulated regions indicate the effect of publication selection. For example, if the distribution is right-skewed, such that relatively large effects are over-represented, this suggests that researchers may be favouring study designs (e.g., not using a multiple baseline design) that offsets a lack of precision by larger estimated effects, enabling their effects to be reported with acceptable levels of statistical significance, and increasing the chances of publication.

Precision can be proxied by sample size (Velickovski & Pugh, 2011). According to sampling theory, larger-sample estimates should be more precise than smaller-sample estimates, with the precision of estimates varying in proportion to the square root of sample size. Adapting this principle to the SCED literature, estimates with a greater number of baseline observations should be more precise than estimates with fewer baseline observations. Reflecting the nature of SCED literature (cf. Kazdin, 2011) we used the number of baseline data-point observations rather than the sample size to proxy precision. Specifically, as SCED research uses small numbers of participants, who act as their own control (i.e., the baseline phase), the square root of the number of baseline observations was used as a proxy measure for precision. A key principle of designing rigorous SCEDs is a stable baseline (Kazdin, 2011). For example, treatment effects can be inflated by a lack of precision at baseline, which is more likely with fewer baseline observations (Ottenbacher, 1986), and are more likely to appear in the published literature, because authors, referees and editors may favour larger effect sizes and/or estimates reported with conventional levels of statistical significance. Accordingly, by comparing the square root of the number of baseline observations with differences in reported effect size across varied baseline observations, we were able to investigate whether the SCED empirical literature reveals traces of publication bias.

Meta-Regression Analysis Modelling Strategy

To apply multivariate meta regression analysis (Stanley & Doucouliagos, 2012) to the SCED literature, we specify the following model to estimate the determinates of our dependent variable, $Effect\ Size_i$ (i.e., the effect sizes reported in the literature):

$$Effect\ Size_i = \hat{\alpha} + \hat{\beta} Sqrt\_Obs_i + \sum_1^k \hat{\lambda}_k MV_{ki} + \varepsilon_i \qquad (1)$$

where $i = 1,\ldots,n$ indexes the n individual estimates reported in the primary literature, ^ signifies a coefficient "to be estimated", and ε_i denotes the usual ordinary least squares regression error term.

The regression analogue of the funnel plot is embedded within this multivariate model. $Sqrt\_Obs_i$ denotes the square root of the number of baseline observations of the i^{th} estimate, which is also measured on the vertical axis of the funnel graph. In the estimated model, the statistical significance of $\hat{\beta}$ indicates the presence of publication bias, while the size gives us a measure of publication bias. In the case of positive publication bias, as indicated by Figure 2, we expect a negative sign. To illustrate, smaller numbers of baseline observations yield imprecisely estimated effects, which favour the selection of larger effects to yield statistically significant effects. Conversely, larger numbers of baseline observations yield more precisely estimated effects, thereby reducing the incentive to favour the reporting of large effects and attenuating publication bias.

In addition, specifying the model with $Sqrt\_Obs_i$ also controls for publication bias. This reflects the nature of regression analysis. Mathematically, each coefficient in a regression model is a partial derivative and so measures the influence of a particular variable on the dependent variable while controlling for the influence of all other variables in the model by holding them constant. In turn, we are able to estimate authentic empirical effects arising from the SCED literature at different values of $Sqrt\_Obs_i$ corresponding to different levels of publication bias, which we anticipate to be potentially large in the presence of a small number of baseline observations but minimal in the presence of a large number of observations.

Sources of heterogeneity in the estimated effect sizes are modelled by the *k* (= 1, …, 10) "moderator variables" (MVs; i.e.; indicator variables with the value of one if the effect size comes from a study with some particular sample or design characteristic and zero otherwise) – where MV_{ki} is the value of the k^{th} moderator variable for the i^{th} effect size in the primary literature, and $\hat{\lambda}_k$ are the effects of each of the *k* moderator variables to be estimated. Table 2 explains the construction of each moderator variable; the mean indicates the proportion of effect sizes associated with the corresponding characteristic. The 10 moderator variables comprise: indicators of the "Design" of each primary study; the "Nature of the outcome variable"; the "Procedural reliability" of the study; "Single versus Group" approach; the type of "Intervention" studied; the "Athlete Standard"; whether the athletes studied are "Adult/Youth"; the "Gender" of the athletes; the "Region" in which the study took place; and "Type of sport" (individual or team).

The estimated regression constant term $\hat{\alpha}$ reflects all systematic influences on the effect size other than the square root of the number of baseline observations (capturing publication bias) and the moderator variables. Accordingly, now that we have explained each element of our model set out in Eq.1, we explain how we use our regression estimates to calculate the "true" or "authentic" empirical *Effect Size* from the literature taking into account:

(i) a range of values of the number of baseline observations $Sqrt\_Obs_i$ (as noted above); and

(ii) that each moderator variable is an intercept shift term, so that the calculation of the range of authentic empirical effects is extended to incorporate the estimated effect of each moderator variable – weighted by mean – on the constant term $\hat{\alpha}$.

Hence, after estimating our model we use the results to calculate a range of "authentic" empirical *Effect Sizes* by substituting: (i) different values of $Sqrt\_Obs_i$ and; (ii) the weighted value of each moderator variable into $\widehat{Effect\ Size} = \hat{\alpha} + \hat{\beta}Sqrt\_obs + \sum_1^k \hat{\lambda}_k(meanMV)_k$, where $\hat{\alpha}$, $\hat{\beta}$, and the $\hat{\lambda}_k$ are obtained from previous estimation of the regression model. The calculations were performed using the *Lincom* command in Stata 15. Moderator variables are binary indicator variables. Hence, for all $Effect\ Size_i$ not associated with a particular source of heterogeneity the corresponding moderator variable is set to zero. Conversely, for $Effect\ Size_i$ that are associated with a particular source of heterogeneity the corresponding moderator variable has value one and the estimated effect $\hat{\lambda}$ is weighted by the mean of the moderator (so that, for example, a moderator associated with 40% of the estimates has twice the weight of one associated with 20%).

As a robustness check we estimated our model both: (i) unweighted (giving each estimate equal weight, regardless of the number of estimates reported by each study); and (ii) weighted by the inverse number of effect sizes reported by the study in which it appears (giving each study equal weight regardless of the number of estimates it reports). In a supplementary file, we include the raw data and syntax we used in Stata (Table 4 includes all short-form variable names to enable replication).

Estimation, Testing Down, and True Effects Procedure. We arrived at our baseline model guided by Ramsey's Regression Equation Specification Error Test (RESET). The main use of the Ramsey test is to detect whether the maintained hypothesis of a linear relationship between the regressors specified by the model is a valid representation of the data (Spanos, 2017). However, it also has power in relation to structural breaks in the data (Darnell, 1994), which may be signalled by the presence of outliers (observations far from the estimated regression line/plane – i.e., with large error terms). Meta regression practitioners are divided with respect to reporting and use of the Ramsey test, although a widely cited set of reporting guidelines contain a general recommendation to pay attention to "Meta regression analysis model specification tests" (Stanley et al., 2013, p. 393). In our study, we interpreted failure of the Ramsey test (by an order of magnitude or more, signified by *p*-values of less than 0.005) as a requirement for 'further investigation' (Darnell, 1994). Overall, our approach proved valuable in identifying: (i) a major structural break in the sample, such as to suggest subsamples arising from two distinct populations; and (ii) a small number of additional outliers.

Interaction Analysis. To complete our empirical analysis, we investigated potential interaction effects between those moderator variables that, across our estimated models, most robustly influence reported effect sizes in the literature.

Results

Publication Bias

The funnel plot (Figure 2) displays the square root of the number of observations in the baseline period (vertical axis) against the effect size (horizontal axis). Studies with a smaller number of observations give the most widely scattered range of effect sizes, while those from studies with a larger number of observations lie within a narrower range, more or less close to the (unweighted) sample mean effect size of 2.92 (SD = 3.80; n = 626; Table 2). To interpret the practical significance of effect sizes, there are numerous guidelines. Cohen suggested a value of 0.20 as small, 0.50 as medium, and 0.80 as large. However, this interpretation is based on group-level, rather than single-case, data. To address this limitation, Parker and Vannest (2009) examined 200 single-system design AB contrasts and suggested the following, more appropriate guidelines: small < 0.87, medium 0.87 to 2.67, and large $>$ 2.67. Accordingly, we can provisionally characterise the representative effect size reported in the SCED literature as "large".

The funnel plot appears right-skewed (a standard test rejects the null of zero skew, $p<0.001$), indicating the presence of positive publication selection bias: specifically, studies with a higher number of observations yield more precise, smaller and tightly clustered effects; while studies with a smaller number of observations yield less precise, larger effects. Therefore, the right-skew may indicate a systematic tendency in the extant literature to over-report large positive effects. Moreover, funnel plots are also used to identify potential outliers (Stanley & Doucouliagos, 2012). Accordingly, the six extreme estimates (ES > 20) lying on the right (positive) side of the plot were identified as outliers and filtered out of all subsequent analyses.

Meta-Regression Results

Table 2 reveals the (unweighted) unconditional means and standard deviations of the variables used in our meta-regression analysis, beginning with effect size. However, given the evidence of positive publication bias in the SCED literature, the unconditional mean effect may be a misleading guide to the true effect. Instead, we use meta regression analysis to gain insight into the size of the "authentic" empirical effect, which – using common meta regression terminology – is the representative effect size estimated "beyond" (i.e., controlling for) both publication bias and sources of heterogeneity.

We identified a structural break between those studies adopting a single-case approach versus a group-based approach. Five hundred and sixty-four observations were from a single case approach and 56 were from a group-based approach. Table 3 indicates that these groups have very different statistical characteristics regarding their effect sizes. Both the unconditional mean values and their standard deviations are substantially different in both the weighted (*M* single-case = 2.59; *SD* = 2.93 vs *M* group-based = 1.65; *SD* = 1.30) and the unweighted samples (*M* single-case = 2.83; *SD* = 3.11 vs *M* group-based = 1.36; *SD* = 1.29).

We concluded that these samples represent different populations and therefore cannot be pooled for meta-regression analysis (to do so, would be to fall into the well-known "apples and oranges" problem). This conclusion is reinforced by our regression analysis: using different model specifications, the pooled sample always fails the Ramsey test by at least an order of magnitude, while regressions for the samples separately reveal satisfactory Ramsey tests. Accordingly, because most of the literature investigates single-case approaches, and thus provides a sample sufficiently large for valid meta-regression analysis, we focus on these for the remainder of our study.

Both our benchmark weighted and unweighted multivariate models include all of our moderator variables. However, in both cases, the Ramsey test is satisfactory at the one per cent level rather than the conventional five per cent level (although this contrasts with the full-sample models, the very best of which fail the Ramsey test by at least an order of magnitude), and there is evidence of extreme multicollinearity. Accordingly, we adopted the standard approach in meta regression studies of "testing down" from the most general model to a specific or parsimonious model that omits irrelevant variables (Stanley & Doucouliagos, 2012). We began by estimating our benchmark weighted and unweighted multivariate models. Then, we removed the variable with the largest standard error (hence, smallest *t*-statistic and largest *p*-value) and re-estimated. This process was continued until all redundant variables were removed. The final models included only variables that are at least close to statistical significance at the 10 per cent level or, in one case (*Int_2*; multiple mental skills in the parsimonious unweighted model), whose retention is necessary for the statistical validity of the model (indicated by a satisfactory Ramsey test).

Using the single-case data, only minimal further data cleaning was necessary to achieve well-specified models. As reported in Table 4: (i) for the unweighted parsimonious model we retained all 564 effects, as the Ramsey test cannot reject this model on grounds of invalid statistical specification; (ii) for the unweighted general model we removed 9 extreme outlier observations from the dataset (retaining $n = 555$) that were revealed as "outer fence residuals" by the "letter value" procedure; (iii) for both the parsimonious and general weighted models we removed 18 outliers according to the "letter value" procedure (hence, $n = 546$ in both cases).

Accordingly, Table 4 reports estimates from four models: two weighted (General and Specific); and two unweighted (General and Specific). In all four models, the variables are jointly significant, indicating a model with explanatory power (in all cases, the p-value on the model F-statistic is less than 0.05). Moreover, in the case of the two parsimonious models, the Ramsey test is satisfactory (in both cases $p > 0.05$) and the multicollinearity apparent in both full models has been eliminated (a mean VIF of less than four or five is generally regarded as satisfactory in this regard), which means that in addition to the model as a whole having explanatory power we can be confident in the separate estimates of the individual effects.

To identify moderator variables as "redundant" to the model, suggests that the respective dimensions of heterogeneity in the literature are not sources of systematically different intervention effects. If we set the bar high, accepting as systemically important only those variables appearing as statistically significant in at least both parsimonious models, then the representative intervention effects identified by our study do not vary systematically by type of intervention (i.e., individual mental skills, multiple mental skills, other), the standard of the participants (i.e., club/recreational, county/regional, collegiate/varsity, professional/international), the gender or gender mix of participants (i.e., female, male, male and female), the particular outcome of the intervention (i.e., psychological, performance, behavioral), the region in which the intervention takes place (i.e., North America, Europe, Australasia), and the age of the participants (i.e., adult, youth). However, we revealed that the following variables do robustly influence the estimates reported in the literature:

1. Type of sport (*Sport_1)*: Interventions with team athletes generated a larger effect size vs individual athletes (1 = team; 0 = individual sport, the omitted category). Three from four estimates are statistically significant (two at the five per cent and one at the 10 per cent level) suggesting a positive influence on estimated effect sizes – other factors held constant – of between 0.99 and 2.15. The fourth estimate is consistent with respect to size but not quite statistically significant.
2. Square root of the number of baseline observations (*SqRt_obs1*): All four estimates are statistically significant (at least at the five per cent level), negative and of similar size – ranging from -0.65 to -0.90. In each case, these estimates indicated substantial positive publication bias (as the number of observations used in studies rises, so the bias is attenuated).
3. Type of design (*Design__multiple_baseline_Yes_1*): In each case, multiple-baseline design (=1; Other= 0, the omitted category) has a negative and highly significant influence on estimated intervention effects (ranging from an average decrease of -1.00 to one of -1.63).
4. Procedural reliability (*Procedural_relability__yes___1_*): In each case, procedural reliability (1= Yes; 0 = No) has a negative and significant influence on estimated intervention effect (ranging from -0.99 to -1.25).

From this analysis, we concluded that the sources of heterogeneity in the effects reported in the SCED literature are less to do with factors beyond the control of researchers (the context of their studies) and more to do with methodological variations that are under their control: studies relying on (i) few baseline observations, and/or (ii) lacking multiple baseline design and/or (iii) lacking procedural reliability will tend to over-estimate effects.

Having identified positive publication bias and the main sources of heterogeneity in the intervention effects reported in the SCED literature, we used our two parsimonious models to calculate the "true" or representative intervention effects revealed by this literature. From the parsimonious weighted model, the representative empirical effect derived from the mean values of each variable and their respective estimated effects was 2.40 ($SD = 2.43$). This is the same as the unconditional mean effect size in the regression sample ($n = 546$), as in theory it must be (Koutsoyiannis, 1977). As such this is just a check on the consistency of our analysis. However, the very high precision of this estimate (t-statistic=12.22 with a p-value < 0.001) yielded a narrow 95% confidence interval of between 2.01 and 2.79, in both cases a substantial effect. Following recalculation at the 25th percentile of the square root of the number of baseline observations (*SqRt_obs1*), publication bias increased and the effect size correspondingly increased – as predicted – to 2.76 ($p < 0.001$) with 95 per cent confidence limits of 2.22 and 3.30. Conversely, at the 75 percentile the effect size decreased to 2.08 ($p < 0.001$) with 95 per cent confidence limits of 1.73 and 2.43. The effect of publication bias is substantial: comparing at the 25th and 75th percentile values of the square root of the number of observations we see a reduction in the estimated effect size of almost a third. Comparing the estimates at the 10th and 90th percentiles yields an even stronger contrast: the confidence intervals are not only similarly narrow but also non-overlapping (at the 10th percentile: 2.38 and 3.87; and at the 90th: 1.05 and 2.07) and the estimated authentic empirical effect size halves, from 3.13 to 1.56. Table 5 sums these results and for comparison adds the equivalent estimates from our unweighted multivariate parsimonious model.

Interaction Analysis Results

For the interaction analyses, we considered the four moderator variables that were significant influences in at least three of the four models (i.e., *Int_2* – multiple mental skills; *Sport_1* – type of sport; *Design__multiple_baseline_Yes_1* – type of design;

Procedural_relability__yes___1 – procedural reliability). We augmented both the preferred weighted and the preferred unweighted parsimonious models with the corresponding interaction terms. In both cases, only the interaction between type of intervention (*Int_2;* multiple mental skills) and design (*Design__multiple_baseline_Yes_1;* type of design) proved to be statistically significant, and only this interaction provided useful information. Although the weighted augmented regression yielded an unsatisfactory Ramsey test ($p = 0.003$), the unweighted regression was satisfactory at the one per cent level ($p = 0.034$) and both were satisfactory with respect to the mean VIF (respectively 1.48 and 1.31). Accordingly, we used Stata's post-estimation *margins* command, applying Bonferroni-adjustment to interpret the interaction effects.

Overall, our parsimonious models with the single significant interaction yielded results consistent with those reported in Table 4. The post-estimation margins calculations suggested that studies with both multiple skills interventions and multiple baselines yielded a reduced effect size compared to studies with: (1) individual skills and other designs (-2.61, $p = 0.001$); (2) individual skills and multiple baseline designs (-2.28, $p < 0.001$); and (3) multiple skills interventions and other designs (-1.90, $p = 0.014$). The other three comparisons were not statistically significant. The results from the unweighted regression are similar. However, because these comparisons are not significantly different from one another, these results provide no evidence that one or another variable is driving (moderating) the influence of the other.

Second, the post-estimation margins calculations supported the implication of these comparisons that the type of intervention (i.e., *Int_2;* multiple mental skills) and study design (i.e., *Design__multiple_baseline_Yes_1*; type of design) exerted their influence independently rather than jointly. In the weighted regression, the marginal effect of multiple baseline design is estimated to be -0.90 ($p = 0.067$) and the marginal effect of multiple skills intervention is -

1.06 ($p = 0.011$); and in the unweighted regression, the marginal effect of multiple baseline design is estimated to be -1.82 ($p < 0.001$) and the marginal effect of multiple skills intervention is -0.67 ($p = 0.122$). These results are in line with the regression results reported in Table 4.

In summary, these post-estimation marginal calculations provided robust evidence that studies with multiple baselines typically report smaller effect sizes; and some evidence that studies of multiple skills interventions likewise typically report smaller effect sizes. The post-estimation calculations also suggest that these effects are independent of one another.

Discussion

The purpose of our study was to extend the review of Barker et al. (2013) by applying meta-regression analyses to address the research question: "Are psychological skills training programmes and behavioral interventions assessed using SCEDs effective in sport?". The findings support previous evidence demonstrating the effectiveness of psychological skills training and behavioral interventions – underpinned by cognitive behavioral principles – in enhancing psychological outcomes, behavior change, and performance (e.g., Brown & Fletcher, 2017; Tod et al., 2011). In addition, our study is the first meta-regression analysis of psychological skills training and behavioral interventions delivered through a SCED framework. In particular, after controlling for typical levels of publication bias in this literature, large increases (i.e., weighted ES = 2.40; unweighted ES = 2.83) in psychological, behavioral, and performance outcomes in studies adopting SCEDs were demonstrated. Accordingly, the findings provide support for: (1) the use of SCEDs to assess psychological, behavior, and performance change in sport (see Barker et al., 2013; Hrycaiko & Martin, 1997; Martin et al., 2004); and (2) the effectiveness of psychological skills training and behavioral interventions – underpinned by cognitive behavioral principles – in sport, thus increasing practitioner confidence in using SCEDs. In other words, methodologically, our study provides unique meta-analytical evidence for SCEDs as an appropriate method in sport to detect meaningful changes in key outcomes, distinguish idiosyncratic effects in response to psychological skills training and behavioral interventions, and assist in the refinement of intervention protocols (see Barker et al., 2011). Theoretically, our study provides support for the application of interventions underpinned by cognitive behavioral principles, which is the popular approach to intervention delivery with athletes (e.g., Hemmings & Holder, 2009).

In addition, our analyses indicated a structural break between SCED studies using a single-case versus a group-based approach. This division was not from a priori theoretical consideration but an emergent finding from the data. Although positive and large, the studies that used a group-based approach brought about lower effect sizes compared to single-case approaches. In general, the group-based studies stated that they adopted SCED principles and conducted analyses on group-level data (e.g., an academy football team) rather than a case-by-case basis. While Kazdin (2011) outlined how SCEDs can be applied to groups, his guidance relates to the application of SCEDs to contexts where between-group evaluations are appropriate. For example, researchers may wish to compare two or more interventions or identify the magnitude of change relative to no treatment (i.e., a control). Further, although the application of SCEDs to a single group with pre and post assessment (e.g., probe design) may enable insights into the assessment of change, this is considered a weak design hindering causal inferences about an intervention. This weakness typically evolves around threats to internal validity not being ruled out. For example, it is possible that in this design participants improve as a function of talking with one another, and therefore show improvements post-intervention. To reduce such threats to interval validity within a single group, continuous assessment through the baseline and intervention phases is recommended to help researchers/practitioners to determine change.

In addition to the general positive effects found, our investigation provides further detail on publication bias and heterogeneous effects – modelled by our moderator variables – in the field of applied sport psychology. First, publication bias is a salient issue in quantitative literatures across the field of psychology (Kühberger et al., 2014) and was evident in our analyses. In our analyses, we found evidence of publication bias in the SCED literature in that the distribution of effect sizes reported by studies using a small number of baseline observations is: (1) widely dispersed; and (2) substantially skewed to the right (i.e., towards overly positive effect sizes). In contrast, studies with a larger number of baseline observations typically reported smaller and more consistent effects.

We estimated authentic empirical effects at different levels of publication bias. We argue, on theoretical grounds, that studies using low numbers of baseline observations and thus reporting results estimated with the lowest levels of precision are the most prone to inflate effect sizes, reflecting a publication bias in the literature. By controlling in our meta regression for the square root of the number of baseline observations reported by each study, we demonstrated that moving from relatively high levels of publication bias (at the 10th percentile) to relatively low levels of publication bias (at the 90th percentile) more than halves the reported effect size. In other words, the authentic empirical effect sizes estimated at typical levels of publication bias (noted above) are likely to be overly optimistic in that they reflect a substantial element of positive publication bias. Accordingly, a more conservative approach would be to take the authentic empirical effects derived from those studies using the largest numbers of baseline observations, thereby reporting the most precise estimates and the least influenced by publication selection bias. In this case, we may characterise the representative effect reported in the SCED literature as “medium” rather than “large” (Parker & Vannest, 2009). Accordingly, we propose that: (a) future SCED investigators consider collecting a larger number of baseline observations (i.e., 8 or more; Ottenbacher, 1986) to reduce publication bias; and (b) journal referees and editors should discriminate according to the quality of the study (e.g., number of baseline observations) and not by the presence of large and/or significant effects. Although it has been acknowledged that it is difficult for sport psychologists to achieve a stable baseline over eight time points (Barker et al., 2013), our meta-regression underscores the importance of doing so. Otherwise, we fall into the trap of reporting inflated effects as a function of not collecting sufficient baseline data. Future SCED researchers must aim to heed these calls.

The use of meta regression procedures in our study demonstrated that the representative effect size varied dependent on the key moderating variables of individual vs team sport, design, and procedural reliability. First, interventions with team sport athletes generated a larger effect size than those with individual sport athletes. It is not clear why, compared to individual sport athletes, team sport athletes reported greater improvements, and this should be a focus for future researchers. Although it is plausible that team sport athletes adhere more closely to psychological skills training and behavioral interventions compared to individual sport athletes, further research is needed to provide clarity on this finding.

Second, compared to multiple-baseline designs, other approaches gave rise to larger intervention effects. Third, studies with no procedural reliability reported significantly larger effect sizes compared to more precise, smaller effects reported in studies with procedural reliability. The adoption of multiple-baseline designs and procedural reliability are central to SCEDs, because they reflect the methodological rigor employed and provide markers of SCED study quality (e.g., Kazdin, 2011). On this point, there are clear implications for applied researchers who should be encouraged to adopt multiple-baseline designs and procedural reliability to ensure that reported intervention effects in peer-reviewed literature are precise and accurate, and less likely to be inflated by methodological shortcomings.

In addition to those noted above, we wish to highlight further implications for researchers in light of our meta regression. The first regards the reporting of appropriate statistical information to allow for calculations of effects (and future meta regression studies). Whereas 59 of the studies identified included sufficient data for effect sizes to be calculated, only 12 studies explicitly reported effect sizes, and a further 50 studies did not report sufficient data for effect size calculation. In other words, applied researchers should heed calls to report effect sizes. Second, the publication bias issue needs to be addressed, and this is not exclusive to SCED literature (see, for example, Kühberger et al., 2014). It may be the case that researchers, reviewers, and/or journal editors are unwilling to submit/accept manuscripts that report non-effects. Instead inspection of the quality and rigor of the study should be considered more important than the results when making publication decisions, a practice that is gaining traction within the social and life sciences (Blanco-Perez & Brodeur, 2019). For instance, in 2015 editors of eight health economics journals published an editorial statement and reminder to referees to accept studies that: "… have potential scientific and publication merit regardless of whether such studies' empirical findings do or do not reject null hypotheses" (Blanco-Perez & Brodeur, p. 1). Following future investigations, the authors concluded that "the editorial statement reduced the extent of publication bias" (Blanco-Perez & Brodeur, p. 27).

Third, researchers are encouraged to use procedural reliability in SCEDs. This is not a new recommendation (see Barker et al., 2013), but is a key principle of SCEDs and our current findings suggest studies that do employ procedural reliability report smaller, more precise effects. Finally, further supporting suggestions from SCEDs researchers (see Barker et al., 2013), a longer baseline period is needed to establish stability and quality. Although potentially difficult in the contextual and ethical constraints of applied research, a sufficient baseline is crucial given that studies with fewer baseline observations produce significantly inflated and less precise effects.

Our study has certain limitations that should be considered when interpreting our findings. First, we were unable to provide a complete picture of the SCEDs literature in sport. As noted above, 50 of the 121 manuscripts (about 41%) did not report sufficient data for effect sizes to be calculated. Despite this, our investigation is the most comprehensive examination of the effectiveness of psychological skills training and behavioral interventions using SCEDs in sport to date. Second, from the characteristics of the included studies, youth athletes, female athletes, and SCED research in cultures beyond western societies are under-represented in the literature. Future researchers should explore these populations and cultures (see Hassmen, Piggot, & Keegan, 2016) when applying psychological skills training and behavioral interventions with SCEDs to enable a more complete picture regarding intervention effectiveness in applied sport psychology. Finally, our data demonstrates the prominence of cognitive-behavioral approaches within applied sport psychology, and therefore, future researchers may wish to consider other approaches (e.g., Acceptance and Commitment Therapy; Hayes, Strosahl, & Wilson, 2012).

In conclusion, our meta-regression analysis of the published literature provides support for the effectiveness of psychological skills training and behavioral interventions in applied sport psychology research, and further demonstrates the large practical effects of implementing SCEDs. On the one hand, supporting the cognitive behavioral approach, this paints a positive picture of the effect of the use of SCED approaches to apply psychological skills training and behavioral interventions with athletes. Yet, the variability documented within the SCED literature appears to be a function of researchers not taking more control of their methodological approaches. For example, studies relying on: (i) few baseline observations and/or (ii) lacking multiple baseline design and/or (iii) lacking procedural reliability will tend to over-estimate effects. Therefore, we conclude that there are key areas of improvement in applied research using SCEDs in sport. Specifically, future researchers should seek to increase the number of baseline observations, use procedural reliability, adopt a multiple baseline design, and report effect size information. Adopting these recommendations will allow for the growth of more rigorous examinations of SCEDs. Moreover, the structural break we found highlights how researchers are adopting SCED principles in their practice with sport teams, and, typically, such work produces still positive, but smaller improvements. Finally, the presence of positive publication bias in the SCED literature points to a need for researchers and those involved in the review process to encourage quality and rigour (rather than reporting positive effects) in the research and publication process. Through these mechanisms, increased understanding of interventions will consequently bolster confidence regarding applied sport psychological services and further delineate insights into effective practice.

*We dedicate our study to the life and work of Professor Aidan Moran who sadly passed away before the manuscript was accepted for publication. Aidan was an inspirational academic and caring friend. May his legacy and influence last forever.

References

*Allen, K. D. (1998). The use of an enhanced simplified habit-reversal procedure to reduce disruptive outburst during athletic performance. *Journal of Applied Behavior Analysis, 31, 489-492.* https://doi.org/10.1901/jaba.1998.31-489

Anderson, A., G., Miles, A., Mahoney, C., & Robinson, P. (2002). Evaluating the effectiveness of applied sport psychology practice: Making the case for a case study approach. *The Sport Psychologist, 16,* 432-453. https://doi.org/10.1123/tsp.16.4.432

*Annesi, J. L. (1998). Applications of the individual zones of optimal functioning model for the multimodal treatment of precompetitive anxiety. *The Sport Psychologist, 12,* 300-316. https://doi.org/10.1123/tsp.12.3.300

*Ashbrook, P. R., Gilham, A., & Barba, D. (2018). Effects of an individualized mental-skills-training program on gold performance: A single-subject research design. *The Sport Psychologist, 32,* 275-289. https://doi.org/10.1123/tsp.2017-0107.

*Barker, J. B., & Jones, M. V. (2005). Using hypnosis to increase self-efficacy: A case study in elite judo. *Sport & Exercise Psychology Review, 1,* 36-42.

*Barker, J. B., & Jones, M. V. (2006). Using hypnosis, technique refinement, and self-modelling to enhance self-efficacy: A case study in cricket. *The Sport Psychologist, 20,* 94-110.

*Barker, J. B., & Jones, M. V. (2008). The effects of hypnosis on self-efficacy, positive and negative affect and sport performance: A case study from professional English soccer. *Journal of Clinical Sport Psychology, 2,* 127-147.

Barker, J. B., McCarthy, P. J., Jones, M. V., & Moran, A. (2011). *Single-case research methods in sport and exercise psychology.* Routledge.

Barker, J. B., Mellalieu, S. D., McCarthy, P. J., Jones, M. V., & Moran, A. (2013). A review of single case research in sport psychology 1997-2012: Research trends and future directions. *Journal of Applied Sport Psychology, 25* (1), 4-32. http://doi.org/10.1080/10413200.2012.709579

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single-case experimental designs: Strategies for studying behavior change* (3rd ed.). Pearson.

*Bell, R. J., Skinner, C. H., & Fisher, L. A. (2009). Decreasing putting yips in accomplished golfers via solution-focused guided imagery: A single-subject research design. *Journal of Applied Sport Psychology, 21,* 1-14. http://doi.org/10.1080/10413200802443776

*Bell, R. J., & Thompson, C. L. (2007). Solution-focused guided imagery for a golfer experiencing the yips: A case study. *Athletic Insight, 9*(1), 52-66.

Blanco-Perez, C. & Brodeur, A. (2019). Publication bias and editorial statement on negative findings. Canadian Centre for Health Economics (CCHE) Working Paper No. 190001 (August): https://www.canadiancentreforhealtheconomics.ca/research/working-papers/

Brown, D. J., & Fletcher, D. (2017). Effects of psychological and psychosocial interventions on sport performance: A meta-analysis. *Sports Medicine*, *47*(1), 77-99. https://doi.org/10.1007/s40279-016-0552-7

Bryan, A. J. (1987). Single-subject designs for evaluation of sport psychology interventions. *The Sport Psychologist, 1,* 283-292.

Burton, D., Naylor, S., & Holliday, B. (2001). Goal setting in sport: Investigating the goal effectiveness paradox. In: R. Singer, H. A. Hausenblas, C. M. Janelle (Eds.), *Handbook of research on sport psychology* (pp. 497–528). Wiley.

*Calmels, C., Berthoumieux, C., & d'Arripe-Longueville, F. (2004). Effects of an imagery training program on selective attention of national softball players. *The Sport Psychologist, 18,* 272-296.

*Callow, N., Hardy, L., & Hall, C. (2001). The effects of a motivational general-mastery imagery intervention on the sport confidence of high-level badminton players. *Research Quarterly for Exercise and Sport*, *72,* 389-400.

*Cunningham, R., & Turner, M. J. (2016). Using Rational Emotive Behavior Therapy (REBT) with Mixed Martial Arts (MMA) athletes to reduce irrational beliefs and increase unconditional self-acceptance. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 34,* 289-309. http://doi/10.1007/s10942-016-0240-4.

Darnell, Adrian C. (1994). *A Dictionary of Econometrics*. Cheltenham, UK: Edward Elgar.

*Deen, S., Turner, M. J., & Wong, R. S. K., (2017). The effects of REBT, and the use of credos, on irrational belies and resilience qualities in athletes. *The Sport Psychologist, 31,* 249-263. doi: https://doi.org/10.1123/tsp.2016-0057

*Didymus, F. F., & Fletcher, D. (2017). Effects of a cognitive-behavioral intervention on field hockey players appraisals of organizational stressors. *Psychology of Sport and Exercise, 30,* 173-185. doi: http://dx.doi.org/10.1016/j.psychsport.2017.03.005

Ellis, A. (1957). Rational psychotherapy and individual psychology. *Journal of Individual Psychology, 13,* 38-44.

*Falls, N., Barker, J. B., & Turner, M. J., (2018). The effects of eye movement desensitization and reprocessing on prospective imagery and anxiety in golfers. *Journal of Applied Sport Psychology, 30,* 171-184, http://doi.org/10.1080/10413200.2017.1345999

*Fitterling, J. M., & Ayllon, T. (1983). Behavioral coaching in classic ballet: Enhancing skill development. *Behavioral Modification, 7*(3), 345-368. http://doi.org/10.1177/01454455830073004.

*Freeman, P., Rees, T., & Hardy, L. (2009). An intervention to increase social support and improve performance. *Journal of Applied Sport Psychology, 21,* 186-200. http://doi.org/10.1080/10413200902785829.

*Galloway, S. M. (2011). The effect of biofeedback on tennis service accuracy. *International Journal of Sport and Exercise Psychology, 9,* 251-266. http://doi.org/10.1080/1612197x.2011.614851

*Galvan, Z. J., & Ward, P. (1998). Effects of public posting on inappropriate on-court behaviors by collegiate tennis players. *The Sport Psychologist, 12,* 419-426.

Gardner, F., & Moore, Z. (2006). *Clinical sport psychology.* Human Kinetics.

Greenspan, M. J., & Feltz, D. L. (1989). Psychological interventions with athletes in competitive situations: A review. *The Sport Psychologist, 3,* 219-236.

*Gregg, M. J., Hrycaiko, D., Mactavish, J. B., & Martin, G. L. (2004). A mental skills package for special Olympic athletes: A preliminary study. *Adapted Physical Activity Quarterly, 21*, 4-18.

*Harwood, C. G., Barker, J. B., & Anderson, R. (2015). Psychosocial development in youth soccer players: Assessing the effectiveness of the 5Cs intervention program. *The Sport Psychologist, 29,* 319-334. http://doi.org/10.1123/tsp.2014-0161.

Hassmén, P., Keegan, R., & Piggott, D. (2016). Rethinking sport and exercise psychology research: Past, present and future. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-48338-6

*Hamilton, R. A., Scott, D., & MacDougall, M. P. (2007). Assessing the effectiveness of self-talk interventions on endurance performance. *Journal of Applied Sport Psychology, 19,* 226-239. http://doi.org/10.1080/10413200701230613

*Hanton, S., & Jones, G. (1999). The effects of a multimodal intervention program on performers: II. Training the butterflies to fly in formation. *The Sport Psychologist, 13,* 22-41.

Hanton, S., & Mellalieu, S.D. (Eds). (2012). *Professional practice in sport psychology: A review*. Routledge.

Hardy, L., & Jones, G. (1994). Current issues and future directions for performance related research in sport psychology. *Journal of Sports Sciences, 12*, 61-92.

Hayes, S. C., Strosahl, K. D., Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change* (2 ed.). Guilford Press.

Hemmings, B., & Holder, T. (2009). *Applied sport psychology: A case-based approach.* Wiley.

*Heyman, S. R. (1987). Research interventions in sport psychology: Issues encountered in working with an amateur boxer. *The Sport Psychologist, 1,* 208-223.

Hrycaiko, D. W., & Martin, G. L. (1996). Applied research studies with single-subject designs: Why so few? *Journal of Applied Sport Psychology, 8,* 183-199.

Ioannidis J. P. A. (2005). Why most published research findings are false. *PLoS Med* 2(8): e124. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/

*Johnson, J. J. M., Hrycaiko, D. W., Johnson, G. V., & Halas, J. M. (2004). Self-talk and female youth soccer performance. *The Sport Psychologist, 18,* 44-59.

*Jones, M. I., Lavallee, D., & Tod, D. (2011). Developing communication and organization skills: The ELITE life skills reflective practice intervention. *The Sport Psychologist, 25,* 35-52.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* Oxford University Press.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press.

*Koehn, S. & Diaz-Ocejo, J. (2016). Imagery intervention to increase flow state: A single-case study with middle-distance runners in the state of Qatar. *International Journal of Sport and Exercise Psychology,* http://doi.org/10.1080/1612197X.2016.1187653

Koutsoyiannis, A. (1977). *Theory of econometrics* (2nd ed). MacMillan Publishing.

*Kladopoulos, C. N., & McComas, J. J. (2001). The effects of form training on foul-shooting performance in members of a women's college basketball team. *Journal of Applied Behavior Analysis, 34*(3), 329-332.

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE 9*(9): e105825. https://doi.org/10.1371/journal.pone.0105825

Kyllo, L. B., & Landers, D. M. (1995). Goal setting in sport and exercise: A research synthesis to resolve the controversy. *Journal of Sport & Exercise Psychology, 17,* 117-137.

*Landin, D., & Herbert, E. P. (1999). The influence of self-talk on the performance of skilled female tennis players. *Journal of Applied Sport Psychology, 11,* 263-282. http://doi.org/10.1080/10413209908404204

*Lauer, L., & Paiement, C. (2009). The playing tough and clean hockey program. *The Sport Psychologist, 23,* 543-561.

Lazarus, R. S., Coyne, J. C., & Folkman, S. (1984). Cognition, emotion and motivation: The doctoring of humpty-dumpty. In K. R. Scherer and P. Ekman (Eds.), *Approaches to Emotion.* (pp. 221-237). London: Lawrence Erlbaum Associates.

*Lerner, B. S., Ostrow, A. C., Yura, M. T., & Etzel, E. F. (1996). The effects of goal-setting and imagery training programs on the free-throw performance of female collegiate basketball players. *The Sport Psychologist, 10*, 382-397.

*Lindsay, P., Maynard, I., & Thomas, O. (2005). Effects of hypnosis on flow states and cycling performance. *The Sport Psychologist, 19,* 164-177.

Mace, R. D. (1990). Cognitive behavioural interventions in sport. In G. Jones and L. Hardy (Eds.), *Stress and performance in sport.* (pp. 203-230). Chichester, England: Wiley.

Mahoney, M. J. (1974). *Cognition and behaviour modification.* Cambridge, Mass: Ballinger.

*Mallett, C. J., & Hanrahan, S. J. (1997). Race modelling: An effective cognitive strategy for the 100m sprinter? *The Sport Psychologist, 11*, 72-85.

*Marlow, C., Bull, S., Heath, B., & Shambrook, C. J. (1998). The use of a single case design to investigate the effect of a pre-performance routine on the water polo penalty shot. *Journal of Science and Medicine in Sport, 1,* 143-155.

Martin G. L., Thomson, K., & Regehr, K. (2004). Studies using single-subject designs in sport psychology: 30 years of research. *The Behavior Analyst, 27,* 263–280.

Martin, G. L., Vause, T., & Schwartzman, L. (2005). Experimental studies of psychological interventions with athletes in competitions: Why so few? *Behavior Modification, 29,* 616-641.

Meichenbaum, D. (1977). *Cognitive behaviour modification: An integrative approach.* New York: Plenum.

Meichenbaum, D. (2010). Cognitive behavioral modification. *Scandinavian Journal of Behaviour Therapy, 6,* 185-192. https://doi.org/10.1080/16506073.1977.9626708

*Mellalieu, S. D., Hanton, S., & Thomas, O. (2009). The effects of a motivational general-arousal imagery intervention upon preperformance symptoms in male rugby union players. *Psychology of Sport and Exercise, 10,* 175-185. http://doi.org/10.1016/j.psychsport.2008.07.003

Meredith, S. J., Dicks, M., Noel, B., & Wagstaff, C. R. D. (2018). A review of behavioral measures and research methodology in sport and exercise psychology. *International Review of Sport and Exercise Psychology, 11,* 25-46. https://doi.org/10.1080/1750984X.2017.1286513

*Messagno, C., Marchant, D., & Morris, T. (2008). A pre-performance routine to alleviate choking in "choking-susceptible" athletes. *The Sport Psychologist, 22,* 439-457.

*McCarthy, P. J., Jones, M. V., Harwood, C. G., & Davenport, L. (2010). Using goal setting to enhance positive affect among junior multi-event athletes. *Journal of Clinical Sport Psychology, 4,* 53-68.

*McKenzie, A. D., & Howe, B. L. (1997). The effect of imagery on self-efficacy for a motor skill. *International Journal of Sport Psychology, 28,* 196-210.

*McKenzie, T. L., & Rushall, B. S. (1974). Effects of self-recording on attendance and performance in a competitive swimming training environment. *Journal of Applied Behavior Analysis, 7*(2), 199-206.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J. & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*(1), 81-95. http://doi.org/10.1007/s12160-013-9486-6

*Milne, D., & Morrison, G. (2015). Cognitive behavioral intervention for the golf yips: A single-case design. *Sport and Exercise Psychology Review, 11,* 20-33.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 6:e1000097. http://doi.org/10.1371/journal.pmed. 1000097

Morgan D. L., & Morgan, R. K. (2009). *Single-case research methods for the behavioral and health sciences.* Sage Publications.

*Munroe-Chandler, K. J., Hall, C. R., Fishburne, G. J., & Shannon, V. (2005). Using cognitive general imagery to improve soccer strategies. *European Journal of Sport Science, 5(1)*, 41-49. http://doi.org/10.1080/17461390500076592

*Neil, R., Hanton, S., & Mellalieu, S. D. (2013). Seeing things in a different light: Assessing the effects of a cognitive-behavioral intervention upon the emotional orientation and performance of golfers. *Journal of Applied Sport Psychology, 25*(1), 106-130. http://doi.org/10.1080/10413200.2012.658901

*Nicholls, A. R., & Polman, R. C. J. (2005). The effects of individualized imagery interventions on golf performance and flow states. *Athletic Insight, 7*(1), 43-66.

Ottenbacher, K. J. (1986). *Evaluating clinical change: Strategies for occupational and physical therapists.* Williams and Wilkins.

*O'Brien, M., Mellalieu, S. D., & Hanton, S. (2009). Goal-setting effects in elite and nonelite boxers. *Journal of Applied Sport Psychology, 21,* 293-306. http://doi.org/10.1080/10413200903030894.

*Pates, J. K., & Cowen, A. (2013). The effect of a hypnosis intervention on performance and flow state of an elite golfer: A single subject design. *International Journal of Golf Science, 2*, 45-53.

*Pates, J. K., & Maynard, I. (2000). Effects of hypnosis on flow states and golf performance. *Perceptual and Motor Skills, 91,* 1057-1075.

*Pates, J. K., Cummings, A., & Maynard, I. (2002). The effects of hypnosis on flow states and three-point shooting performance in basketball players. *The Sport Psychologist, 16,* 34-47.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research:

Nonoverlap of all pairs. *Behavior Therapy, 40*(4), 357-367. http://dx.doi.org/10.1016/j.beth.2008.10.006

*Ploszay, A. J., Gentner, N. B., Sinner, C. H., & Wrisberg, C. A. (2006). The effects of multisensory imagery in conjunction with physical movement rehearsal of golf putting performance. *Journal of Behavioral Education, 15,* 249-257. http://doi.org/10.1007/s10864-006-9034-6

*Polaha, J., Allen, K., & Studley, B. (2004). Self-monitoring as an intervention to decrease swimmers' stroke counts. *Behavior Modification, 28,* 261-275. http://doi.org/10.1177/0145445503259280

*Post, P., Muncie, S., & Simpson, D. (2012). The effects of imagery training on swimming performance: An applied investigation. *Journal of Applied Sport Psychology, 24*(3), 323-337. http://doi.org/10.1080/10413200.2011.643442

*Robazza, C., Pellizzari, M., & Hanin, Y. (2004). Emotion self-regulation and athletic performance: An application of the IZOF model. *Psychology of Sport and Exercise, 5,* 379-404.

Rumbold, J., Fletcher, D., & Daniels, K. (2012). A systematic review of stress management interventions with sport performers. *Sport, Exercise, and Performance Psychology, 1*(3), 173-193.

*Rush, D. B., & Ayllon, T. (1984). Peer behavioral coaching: Soccer. *Journal of Sport Psychology, 6,* 325-334.

*Scott, D., Scott, L. M., & Howe, B. L. (1998). Training anticipation for intermediate tennis players. *Behavior Modification, 22,* 243-261. http://doi.org/10.1177/01454455980223002

*Scott, L. M., Scott, D., Bedic, S. P., & Dowd, J. (1999). The effect of associative and dissociative strategies on rowing ergometer performance. *The Sport Psychologist, 13,* 57-68.

*Shambrook, C. J., & Bull, S. J. (1996). The use of a single-case research design to investigate the efficacy of imagery training. *Journal of Applied Sport Psychology, 8*(1), 27-43.

*Shapiro, E. S., & Shapiro, S. (1985). Behavioral coaching in the development of skills in track. *Behavior Modification, 9*(2), 211-224. http://doi.org/10.1177/01454455850092005.

*Shearer, D. A., Mellalieu, S. D., Shearer, C., & Roderique-Davies, G. (2009). The effects of a video-aided imagery intervention upon collective efficacy in an international Paralympic wheelchair basketball team. *Journal of Imagery Research in Sport and Physical Activity, 4*(1), 1-25

*Silva III, J. M. (1982). Competitive sport environments: Performance enhancement through cognitive intervention. *Behavior Modification, 6*(4), 443-463. http://doi.org/10.1177/01454455820064001.

Smith, R. E. (1980). A cognitive affective approach to stress management for athletes, in C. A. Nadeau, W. R. Halliwell, K. M. Newell and G. C. Roberts (Eds.). *Psychology of motor behaviour and sport.* Champaign, Illinois: Human Kinetics.

Smith, R. E. (1989). Applied sport psychology in an age of accountability. *Journal of Applied Sport Psychology, 1*(2), 166-180.

Smith, J. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods, 17*(4), 510-550.

*Smith, S. L., & Ward, P. (2006). Behavioral interventions to improve performance in collegiate football. *Journal of Applied Behavior Analysis, 39*(3), 385-391.

Spanos, A., (2017). Mis-specification testing in retrospect. *Journal of Economic Surveys*. On-line, prepublication: http://doi.org/10.1111/joes.12200.

Stanley, T.D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*, 309–345.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, *70*(1), 103-127.

Stanley, T.D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. Routledge.

Stanley, T.D., Doucouliagos, H., Giles, M., Heckemeyer, J.H., Johnston, R., Laroche, P., Nelson, J., Paldam, M., Poot, J., Pugh, G., Rosenberger, R., & Rost, K. (2013). Meta-analysis of economics research: Reporting guidelines. *Journal of Economic Surveys*. *27*(2), 390–394.

Suinn, R. (1972). Behaviour rehearsal training for ski racers: *Behavior Therapy, 3***,** 519-520.

*Swain, A., & Jones, G. (1995). Effects of goal-setting interventions on selected basketball skills: A single-subject design. *Research Quarterly for Exercise and Sport, 66*(1), 51-63.

*Swainston, S., Gentner, N., Biber, D., Czech, D. R., Joyner, B., & Easton, L. E. (2012). The effect of PETTLEP imagery in a pre-shot routine on full swing golf shot accuracy: A single subject design. *International Journal of Golf Science, 1*, 140-163.

*Thelwell, R. C., & Maynard, I. (2003). The effects of a mental skills package on 'repeatable good performance' in cricketers. *Psychology of Sport and Exercise, 4,* 377-396.

*Thomas, O., Maynard, I., & Hanton, S. (2007). Intervening with athletes during the time leading up to competition: Theory to practice II. *Journal of Applied Sport Psychology, 19,* 398-418.

Tod, D., Hardy, J., & Oliver, E. (2011). Effects of self-talk: A systematic review. *Journal of*

Sport & Exercise Psychology, 33, 666-687.

*Turner, M., & Barker, J. B. (2013). Examining the efficacy of Rational-Emotive Behavior Therapy (REBT) on irrational beliefs and anxiety in elite youth cricketers. *Journal of Applied Sport Psychology, 25*(1), 131-147. http://doi.org/10.1080/10413200.2011.574311.

*Turner, M. J., Ewen, D., & Barker, J. B. (2018). An idiographic single-case study examining the use of Rational Emotive Behavior Therapy (REBT) with three amateur golfers to alleviate social anxiety. *Journal of Applied Sport Psychology.* http://doi.org/10.1080/10413200.2018.1496186

*Turner, M. J., & Davis, H. S. (2019). Exploring the effects of Rational Emotive Behavior Therapy on the irrational beliefs and self-determined motivation of triathletes. *Journal of Applied Sport Psychology, 31,* 253-272. https://doi.org/10.1080/10413200.2018.1446472

*Turner, M. J., Slater, M. J., & Barker, J. B. (2014). The season-long effects of rational emotive behavior therapy on the irrational beliefs of professional academy soccer athletes. *International Journal of Sport Psychology, 44,* 429–451.

Vealey, R. S. (1994). Current status and prominent issues in sport psychology interventions. *Medicine and Sciences in Sports and Exercise, 26,* 495-502.

Velickovski, I., & Pugh, G. (2011). Constraints on exchange rate flexibility in transition economies: A meta-regression analysis of exchange rate pass-through. *Applied Economics*, *43*, 4111–4125.

*Vertopoulos, E., & Turner, M. J. (2017). Examining the effectiveness of a Rational Emotive Personal-Disclosure Mutual-Sharing (REPDMS) intervention on the irrational beliefs and rational beliefs of the Greek adolescent athletes. *The Sport Psychologist, 31,* 264-274. https://doi.org/10.1123/tsp.2016-0071

*Wakefield, C., & Smith, D. (2011). From strength to strength: A single-case design study of PETTLEP imagery frequency. *The Sport Psychologist, 25,* 305-320.

*Ward, P., & Carnes, M. (2002). Effects of posting self-set goals on collegiate football players' skill execution during practices and games. *Journal of Applied Behavior Analysis, 35,* 1-12.

Weinberg, R. S., (2019). Psychological skills training. In D. Hackfort, R. J. Schinke & B. Strauss (Eds.), *Dictionary of sport psychology: Sport, exercise and performing Arts* (pp. 230-231). Elsevier. https://doi.org/10.1016/C2016-0-03935-9

Wessler, R. L. (1986). Conceptualizing cognition's in the cognitive-behavioural therapies. In W. Dryden & W. L. Golden (Eds.), *Cognitive-Behavioural approaches to psychotherapy* (pp. 1-30). London. Harper and Row

*Wood, A. G., Barker J. B., Turner, M. J., & Sheffield, D. (2018). Examining the effects of rational emotive behavior therapy on performance outcomes in elite Paralympic athletes. *Scandinavian Journal of Medicine and Science in Sports, 28,* 329-339. https://doi.org/10.1111/sms.12926

*Ziegler, S. G. (1994). The effects of attentional shift training on the execution of soccer skills: A preliminary investigation. *Journal of Applied Behavior Analysis, 27*(3), 545-552. https://doi.org/10.1901/jaba.1994.27-545

Key: *Indicates that the study was included in the meta regression analysis

Figure 1

PRISMA flow diagram detailing the research study identification and selection process (Moher et al., 2009)

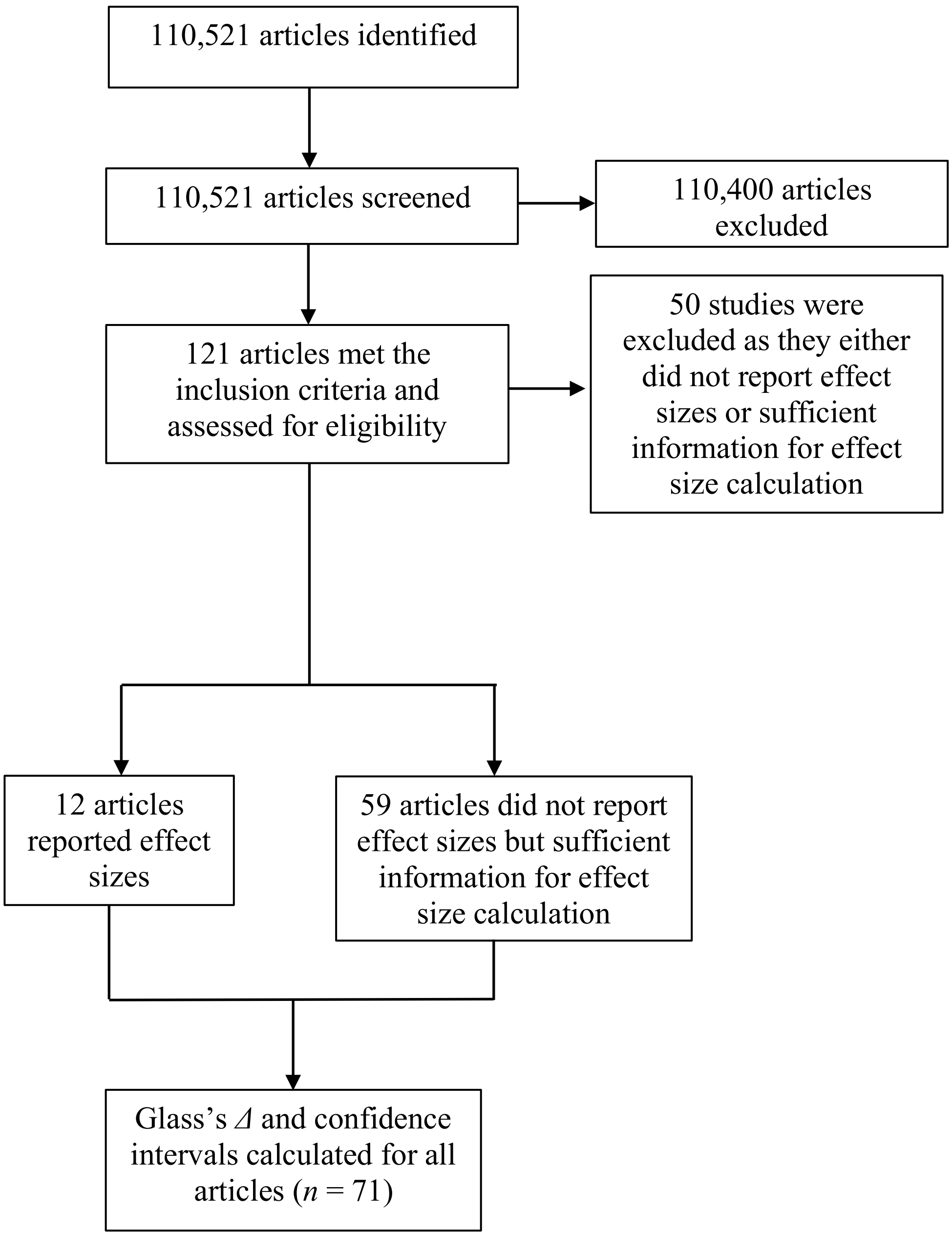

Figure 2

Funnel plot displaying transformed effect size by the square root of the number of baseline observations (Mean effect size = 2.92, indicated by the vertical red line; n = 626)

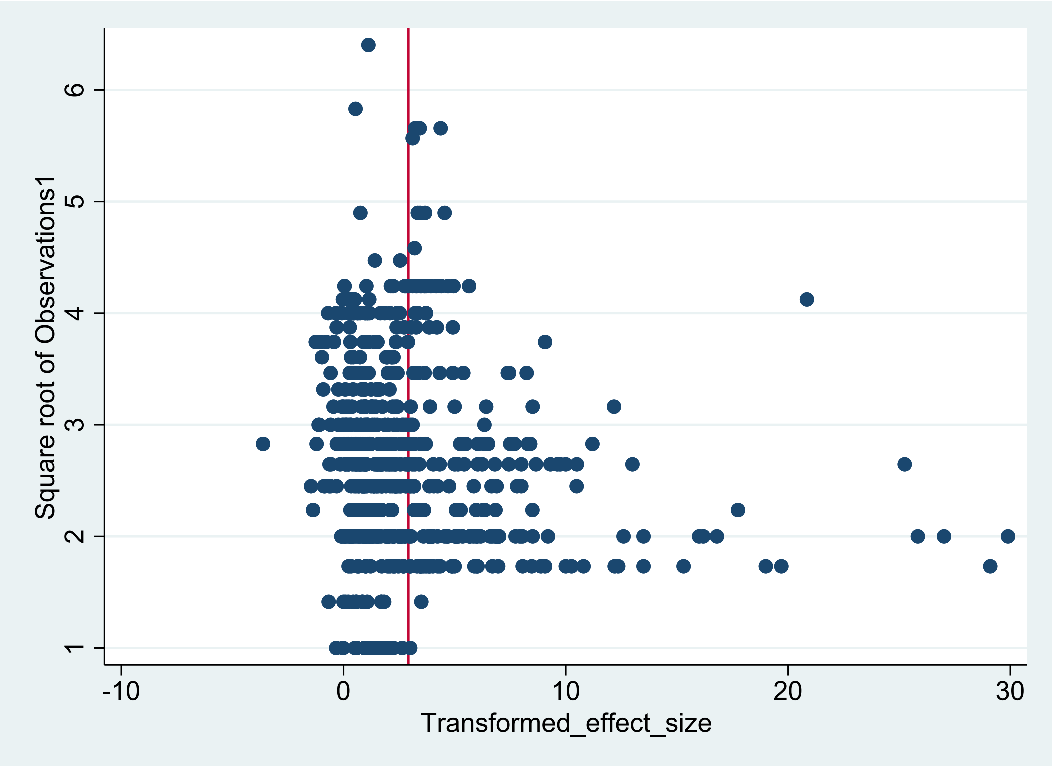

Table 1

Study characteristics of the 71 articles

Characteristic	Studies, N (%)
Region	
North America	31 (43.66)
Europe	37 (52.11)
Australia	3 (4.23)
Intervention*	
Individual mental skill	36 (50.00)
Multiple mental skill	23 (31.94)
Other	13 (18.06)
Design	
Multiple baseline	54 (76.06)
Other	17 (23.94)
Procedural Reliability	
Yes	36 (50.70)
No	35 (49.30)
Sport	
Individual	44 (61.97)
Team	25 (35.21)
Individual and team	2 (2.82)
Standard*	
Recreational/club	23 (31.94)
County/Regional	17 (23.61)
Collegiate	15 (20.83)
Professional/international	17 (23.61)
Participant	
Adult	52 (73.24)
Youth	19 (26.76)
Gender	
Male	42 (59.15)
Female	14 (19.72)
Male and female	14 (19.72)
Not reported	1 (1.41)
Outcome*	
Performance	35 (42.17)
Psychological	36 (43.37)
Behavioral	12 (14.46)
Total *N*	71 (100)

* the total of these variables equal more than 71 because either: (1) one study included more than one intervention (i.e., Lerner et al, 1996); or (2) athlete standard (i.e., O'Brien, Mellalieu, & Hanton, 2009); or (3) multiple studies included more than one type of outcome variable (e.g., Barker & Jones, 2006).

Table 2

Descriptive statistics for the Meta-Regression Analyses models

Variable	Variable Name in Stata	Omitted category	Unweighted Mean (SD)	Weighted Mean (SD)	Min / Max
Effect size	Transformed_effect_size	N/A	2.69 (3.02)	2.47 (2.80)	-3.62 / 19.70
Square root baseline observations	SqRt_obs1	N/A	2.68 (.91)	2.64 (.88)	1 / 6.40
Design	Design__multiple_baseline_Yes_1	Other (= 0)	.80 (.40)	.78 (.42)	0 / 1
Intervention					
Multiple mental skills	Int_2	Individual mental skill (=0)	.42 (.49)	.35 (.48)	0 / 1
Other	Int_3	Individual mental skill (=0)	.17 (.38)	.17 (.37)	0 / 1
Athlete standard					
County/regional	Standard_2	Club/recreational (= 0)	.29 (.45)	.21 (.41)	0 / 1
Collegiate/varsity	Standard_3	Club/recreational (= 0)	.22 (.42)	.26 (.44)	0 / 1
Professional/international	Standard_4	Club/recreational (= 0)	.23 (.42)	.24 (.43)	0 / 1
Adult/youth	Participants__adult___1__youth_	Youth (= 0)	.80 (.40)	.74 (.44)	0 / 1
Nature of outcome variable					
Performance	Outcome_1	Psychological (= 2)	.34 (.48)	.43 (.50)	0 / 1
Behavioral	Outcome_3	Psychological (= 2)	.08 (.27)	.14 (.34)	0 / 1
Region					
Europe	Region_1	North America (= 0)	.36 (.48)	.42 (.49)	0 / 1
Australasia	Region_3	North America (= 0)	.08 (.27)	.07 (.25)	0 / 1

Procedural reliability	Procedural_relability__yes___1_	No (= 0)	.49 (.50)	.48 (.50)	0 / 1
Gender					
Female	Gender_1	Male (= 1)	.27 (.44)	.19 (.40)	0 / 1
Mixed	Gender_3	Male (= 1)	.16 (.37)	.20 (.40)	0 / 1
Type of sport	Sport_1	Individual (= 0)	.59 (.52)	.65 (.54)	0 / 1
Single vs Group Approach	Presented_data_DV	Single-case (= 1)	1.09 (.29)	1.13 (.33)	1 / 2

Note. Moderators: Design (multiple-baseline, other); Intervention (individual mental skill, multiple mental skills, other); Standard (club/recreational, county/regional, collegiate/varsity, professional/international); Participant (adult, youth); Outcome (psychological, performance, behavioral); Region (North American, Europe, Australasia); Procedural reliability (yes, no); Gender (female, male, male and female); Sport (team, individual); Approach (single case, group).

Table 3
Mean and standard deviation of the effect sizes from single-case and group-based approaches

	Unweighted Sample	Weighted Sample
Group-based approach		
Mean	1.36	1.65
Standard Deviation	1.29	1.30
Single-case approach		
Mean	2.83	2.59
Standard Deviation	3.11	2.93
H_0: Equal Variance	p=0.0000	p=0.0000
H_0: Equal Mean	p=0.0000	p=0.0000

Table 4

Multivariate meta regressions, unweighted and weighted OLS estimates

Dependent variable: Transformed effect size - glass's delta (TRANSFORMEDEFFECTSIZEGLASSS)

Cluster-robust standard errors (adjusted for 60 clusters in AUTHOR) used to compute t-statistics

		Weighted								**Unweighted**							
		Full ("general") model				Parsimonious ("specific") model				Full ("general") model				Parsimonious ("specific") model			
		Number of obs = 546 *F*(18, 59) = 1.95 Prob > F = 0.029 R^2 = 0.16				Number of obs = 546 *F*(7, 59) = 2.79 Prob > F = 0.014 R^2 = 0.14				Number of obs = 555 *F*(18, 59) = 8.32 Prob > F = 0.0000 R^2 = 0.24				Number of obs = 564 *F*(5, 60) = 6.22 Prob > F = 0.0001 R^2 = 0.19			
Variable (omitted)	**Name in Stata (category)**	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**Coef.**	**Std. Err.**	**t**	**P>|t|**
Square root of baseline observations	SqRt_obs1	-0.65	0.28	-2.33	0.023	-0.73	0.24	-3.05	0.003	-0.83	0.27	-3.10	0.003	-0.90	0.27	-3.30	0.002
Intervention (individual)	Int_2 (multiple mental skill)	-1.28	0.47	-2.74	0.008	-1.08	0.41	-2.61	0.011	-0.93	0.47	-1.97	0.054	-0.63	0.45	-1.41	0.164
Intervention (individual)	Int_3 (other intervention)	-0.01	0.60	-0.01	0.99					0.25	0.59	0.43	0.667				
Design (other)	Design__multiple_baseline_Yes_1 (multiple baseline)	-1.21	0.52	-2.34	0.023	-1.00	0.49	-2.03	0.047	-1.61	0.68	-2.38	0.021	-1.63	0.54	-3.04	0.003
Team vs individual sport (individual)	Sport_1 (team)	0.95	0.59	1.62	0.111	0.99	0.54	1.83	0.072	1.98	0.61	3.26	0.002	2.15	0.60	3.61	0.001
Athlete standard (club/recreational)	Standard_2 (county/regional)	0.12	0.64	0.19	0.851					0.58	0.84	0.69	0.49				
Athlete standard (club/recreational)	Standard_3 (collegiate/varsity)	0.02	0.60	0.04	0.968					0.77	0.75	1.03	0.308				
Athlete standard (club/recreational)	Standard_4 (professional/international)	0.04	0.60	0.06	0.953					0.35	0.85	0.42	0.677				
Gender (male)	Gender_1 (female)	-0.22	0.60	-0.36	0.72					0.02	0.59	0.03	0.977				
Gender (male)	Gender_3 (mixed)	0.73	0.46	1.58	0.119	0.69	0.43	1.61	0.112	0.87	0.40	2.16	0.035				
Outcome (psychological)	Outcome_1 (performance)	-0.47	0.63	-0.75	0.459					-0.33	0.76	-0.43	0.667				
Outcome (psychological)	Outcome_3 (behavioral)	-0.67	0.81	-0.83	0.411					0.33	0.83	0.40	0.69				
Region (North America)	Region_1 (Europe)	-0.20	0.71	-0.28	0.784					-0.66	0.95	-0.70	0.489				

Region (North America)	Region_3 (Australasia)	-0.98	0.62	-1.58	0.12	-0.72	0.43	-1.66	0.101	-1.30	0.73	-1.79	0.078				
Athlete age (youth)	Participants_adult__1_youth_ (adult)	0.13	0.50	0.26	0.794					-0.64	0.99	-0.65	0.519				
Procedural reliability (No)	Procedural_reliability_yes__1_ (Yes)	-0.99	0.58	-1.71	0.092	-1.19	0.50	-2.39	0.02	-1.25	0.51	-2.46	0.017	-1.09	0.44	-2.46	0.017
Years	Year	-10.16	8.73	-1.16	0.25					-14.16	10.41	-1.36	0.179				
Years squared	Years_Sq	0.003	0.002	1.16	0.251					0.004	0.003	1.36	0.179				
Constant	_cons	10200.29	8741.04	1.17	0.248	5.18	0.92	5.61	0.000	14199.42	10424.83	1.36	0.178	5.49	0.87	6.33	0.000
		Ramsey RESET test H0: model has no omitted variables. *F*(3, 524) = 3.61 Prob > F = 0.013 Mean VIF: 25634.71				**Ramsey RESET test** H0: model has no omitted variables. *F*(3, 535) = 2.31 Prob > F = 0.075 Mean VIF: 1.20				**Ramsey RESET test** H0: model has no omitted variables. *F*(3, 533) = 3.53 Prob > F = 0.0148 Mean VIF: 30771.34				**Ramsey RESET test** H0: model has no omitted variables. *F*(3, 555) = 2.49 Prob > F = 0.0596 Mean VIF: 1.09			

Note: Those highlighted in grey scale are significant at $p < .10$

Table 5

Authentic effect sizes from the parsimonious models

	Weighted	Unweighted
Square root of observations, percentile		
10	3.13***	3.72***
25	2.76***	3.48***
Mean	2.40***	2.83***
75	2.08***	2.44***
90	1.56***	1.69***

Note. *** denotes $p < 0.01$ (i.e., statistically significant at the one per cent level)