WHAT CAN WE KNOW ABOUT THE EFFECTIVENESS OF COACHING? A META-ANALYSIS BASED ONLY ON RANDOMIZED CONTROLLED TRIALS

ERIK DE HAAN
Hult International Business School
VU University Amsterdam

VIKTOR O. NILSSON
Hult International Business School
Norwegian School of Economics

The study involved a comprehensive meta-analysis of 37 randomized controlled trial (RCT) studies of workplace and executive coaching programs written in the English language between 1994 and 2021, using rigorous inclusion criteria designed to align with the robustness of statistical significance and coaching applications. Analysis of 39 coaching RCT samples, with a total sample size of $n = 2,528$, yielded a statistically significant effect of workplace coaching across all leadership and personal outcomes. Our best estimate of a standard effect size for coaching of $g = .59$ fell well within the moderate range, although there were indications of significant publication bias, as expected. Furthermore, we develop a model that is based on the leverage from coregulation between coach and coachee. The model was confirmed in five different predictions: moderation analyses indicated that effects are larger with self-reported outcomes, as distinct from observed outcomes; they are larger with qualified coaches in nonleadership applications; and they are somewhat larger with female coachees, while they do not depend much on the length of the overall coaching assignment. Taken together, these findings provide clear and new evidence of the efficacy of coaching interventions in a variety of applications.

Coaching is a customized personal development journey that promotes the client's decision-making and performance through conversations, making use of shared inquiry, reflection, support, and challenge. As a pathway for personal and professional growth, coaching is now firmly established in the sports and business worlds and enjoys a growing popularity in the teaching and medical professions. In response to the coaching profession's growth over the last two decades, research into workplace coaching outcomes has increased significantly, and has produced a wide and varied evidence base for its effectiveness. Of the 160 unique and rigorous empirical studies summarized in De Haan (2021), 11 were published in 2019 and 14 in 2020 alone.

Despite a growing base of empirical research, the effectiveness of workplace coaching has not been established beyond doubt. Most published studies show clear indications of effectiveness, although there are

several counterexamples and the case against coaching's effectiveness can still be made, especially when publication biases are estimated (de Haan, 2021). Beyond establishing general effectiveness, there are clear questions around how many sessions are needed to achieve a satisfactory level of outcomes. As this study will show, a meta-analysis can now establish not only a convincing measure of general effectiveness, but also the effects of dosage (session numbers), coachee gender, coach and coachee job roles, and even the types of outcomes measured. Eventually, one would expect to show differential effects for coaching models and years of experience, although most researchers expect common factors to dominate, which means that interventions, qualifications, and experience levels are not expected to be very distinctive when regressed against outcomes.

There have been four previously published, explorational meta-analyses of coaching effectiveness (Burt & Talati, 2017; Jones, Woods, & Guillaume, 2015; Sonesh, Coultas, Lacerenza, Marlow, Benishek, & Salas, 2015; Theeboom, Beersma, & van Vianen, 2014), but they were all conducted before a recent

Corresponding author: Erik de Haan (erik.dehaan@ ashridge.org.uk).

growth of randomized controlled trials (RCTs). They encompass 18, 24, 17, and 11 studies, respectively, and miss out on some of the earlier RCTs. We believe each of those analyses has insufficient power to demonstrate small effects or reject the null hypothesis—see, for example, our power analysis in the Methodology section. Moreover, they are all based on study selections that include not just RCTs but also quasi-experimental designs wherein the control group was not randomized, which may have overstated the effect sizes. Only one of the earlier meta-analyses, Jones et al. (2015), reported a "prediction interval" (PI), which is meta-analysis's most powerful measure of the significance of overall effectiveness, and that study only reported a nonstandard 80% PI.

The most reliable research protocol, the randomized controlled trial, has grown considerably in coaching. RCT applications have more than tripled since the most recent overview by Burt and Talati in 2017. Considering this strong growth in research, helped by a growth in PhDs in coaching and a growing number of medical applications of workplace coaching, we believe it is time to attempt a new meta-analysis study, this time exclusively based on RCTs, so that artificial biases are avoided, and effect sizes are more comparable. In a meta-analysis of studies with between-group designs, effect sizes are always measured against those of a control group in the same circumstances, which substantially reduces same-source and selection biases, Hawthorne effects, and other false positives. For this meta-analysis, we have brought together all 37 RCT studies in workplace, executive, and life coaching, containing 40 independent samples. Where data were incomplete, we contacted the authors. For one article, the authors could no longer report the full data, so our meta-analysis is based on 36 rigorous RCT studies, which were published between 1994 to 2021, with 26% of the included samples published since the latest published meta-analysis (Burt & Talati, 2017)

Workplace (including leadership or executive) coaching has characteristics that lend it very well to outcome research. Coaching is a goal-directed activity that mostly takes place during carefully time-bounded, discrete sessions in a neutral, delimited location. This creates something similar to a "lab" situation with clearly described, anticipated outcomes, where all changes can be attributed to a precise number of hours and sessions, and to those sessions alone. There are also challenges in conducting coaching research, mostly to do with the commissioning of coaching programs, where organizational sponsors want to determine who gets access and when, which makes the creation of an equal and randomized control group precarious in realistic situations (see, e.g., the disrupted trials of Egan & Song, 2005, and Ungerer, Heinzelmann, Baltes, & König, 2019).

We believe nevertheless that, with careful formulation of theory and hypotheses, a new meta-analysis can give insight into which implementations of coaching can be most (cost-) effective and can resolve challenges in the implementation of leadership coaching, such as the number of sessions or qualification levels of coaches. This meta-analysis can also test if coachee self-reported outcomes are worth measuring even if they are notoriously biased toward false positives (Grover & Furnham, 2016). It would be able to show that the findings from less rigorous but, in many cases, more realistic trials still support similar predictions as the more rigorous studies.

In summary, this study was undertaken to (a) provide a clear PI for effectiveness of workplace coaching based on the relatively less biased RCT studies, (b) fill gaps and update the current systematic review literature, and (c) identify clear and significant moderators of this effectiveness, by studying differentials in the studies related to type of coaches, coachees, and outcome measures.

THEORETICAL MODEL UNDERPINNING OUR META-ANALYSIS

Based on results in psychotherapy (Cuijpers, Karyotaki, De Wit, & Ebert, 2020) and mentoring (Stoeger, Balestrini, & Ziegler, 2021), and from a glance through the results from RCTs in coaching, we would expect to be able to demonstrate a moderate level of effectiveness and a PI that does not include zero. We base all our other hypotheses on the cocreated nature of coaching (Erdös & Ramseyer, 2021; Wasylyshyn, 2022), where coregulation (Sbarra & Hazan, 2008), relational attunement (Yip, Trainor, Black, Soto-Torres, & Reichard, 2020), and nonverbal synchrony (Koole & Tschacher, 2016) are significantly shaping measurable outcome factors. Synchrony, responsiveness, and coregulation within sessions have been shown to relate to better therapy and coaching outcomes (e.g., Ianiro, Lehmann-Willenbrock, & Kauffeld, 2015; Ramseyer & Tschacher, 2011; Watson & Wiseman, 2021). The terms "responsiveness," "relational attunement," "cocreation," and "coregulation" refer to the ability of coach and coachee to respond flexibly and constructively to new events—that is, explicitly processing and addressing new information in the present moment.

Nonverbal (interpersonal) synchrony can be seen as evidence for this coregulation: it is defined by synchronized (and slightly delayed) changes within coach and coachee—for example, vocal pitch, bodily movements, and even their physiological processes.

Coregulation boils down to the ability to influence one another in a "relational dance" (De Haan, 2008), moment by moment, but also longer term. One would expect forces from (self-)motivation and status through to positional power and relational leverage to play a role in shaping the willingness and ability to cocreate. Indeed, abundant evidence has already been found for the contribution of such leverage-underpinning coregulation to predict coaching effectiveness: (a) the credibility (e.g., Boyce, Jackson, & Neal, 2010; Bozer, Sarros, & Santora, 2014; Grant, 2014; Sue-Chan & Latham, 2004) and dominant-friendliness (e.g., Ianiro & Kauffeld, 2014; Ianiro et al., 2015; Ianiro, Schermuly, & Kauffeld, 2013) of the coach; (b) the intrinsic motivation of the coachee (e.g., Bozer, Sarros, & Santora, 2013); and (c) the trust within or strength of the coaching alliance (e.g., Boyce et al., 2010; De Haan, Grant, Burger, & Eriksson, 2016; Kim & Kuo, 2015; Zimmermann & Antoni, 2020). Similar coregulated factors have been shown to be important in the effectiveness of psychotherapy (e.g., Ramseyer & Tschacher, 2011; Stiles, Honos-Webb, & Surko, 1998; Wieder & Wiltshire, 2020).

We would therefore hypothesize that *leverage* as expressed by coregulation—that is, by influence, credibility, motivation, and trust within the coaching relationship—is bound to be the determining variable for the effectiveness of coaching. Our model as illustrated in Figure 1 predicts that coaching impact correlates with the *leverage* that the coach and coachee have on one another—the ability of both partners to influence each other and to shape a coaching relationship that is safe, independent, and conducive to learning and change.

This Study's Hypotheses

This study's hypotheses flow from our model, which takes the *leverage* within the cocreated and responsively regulated coaching relationship as the fundamental variable that predicts outcomes (see Figure 1 for an illustration of the hypotheses).

Overall effect size. The most rigorous and helpful estimate of the range of effect sizes to be expected is the PI. Of the previous meta-analyses, only Jones et al. (2015) reported an 80% PI in their meta-analysis study, which ran from −0.16 to 0.97 for corrected effect sizes. We expect, with more samples, to substantially improve on this interval. We will report the standard 95% PI, which is a lot more stringent, and we hypothesize to find it entirely above zero, which would produce a strong indication for the effectiveness of coaching. Therefore:

Hypothesis 1. Coaching will demonstrate positive and moderate effects that can be shown to be significantly different from zero.

The outcomes of coaching. The aim of coaching and the measures of coaching outcomes varies between several studies. Evaluating the different outcomes of coaching research is therefore important. A systematic review of workplace coaching outcomes comprising 110 peer-reviewed studies by Athanasopoulou and Dopson (2018) suggested that there is evidence of three categories of coaching outcome—namely, personal development, behavioral changes, and the coachee and their work. The personal development category includes specific outcomes such as health-related measures (reduced stress, increased life satisfaction, and experienced support). The behavioral change outcome includes specific skills and performance measures expressed by better communication skills, management skills, and team-building skills. Lastly, the "coachee and their work" category includes outcomes focusing on psychological preparedness that affect workplace productivity, such as self-awareness, psychological capital, and self-efficacy. There is also support in the literature for a goal attainment approach in coaching, where the coach strives to achieve the intended outcomes by working on the coachee's goal (Grant, 2012). Workplace coaching with a goal attainment focus does not always measure improved health benefits, workplace skills, or readiness.

We would expect the strength of coaching outcomes to be dependent on the methodology of data collection. We expect effect sizes to be higher for self-scores than for scores by others and for objective ratings simply because coach and coachee have more leverage on the same-source biases (see De Haan, 2021, for an extensive comparison of self-scores, multisource feedback scores, and more objective and independent outcome measures). Given the variety of coaching outcomes within the review by Athanasopoulou and Dopson (2018), and the support for goal-attainment coaching approaches (Grant, 2012), we posit:

Hypothesis 2a. Coaching will have positive effects for well-being, workplace skills, preparedness, and goal attainment.

Hypothesis 2b. Effect sizes will be skewed upwards for self-scores.

FIGURE 1
Overview of the Four Hypotheses in the Study
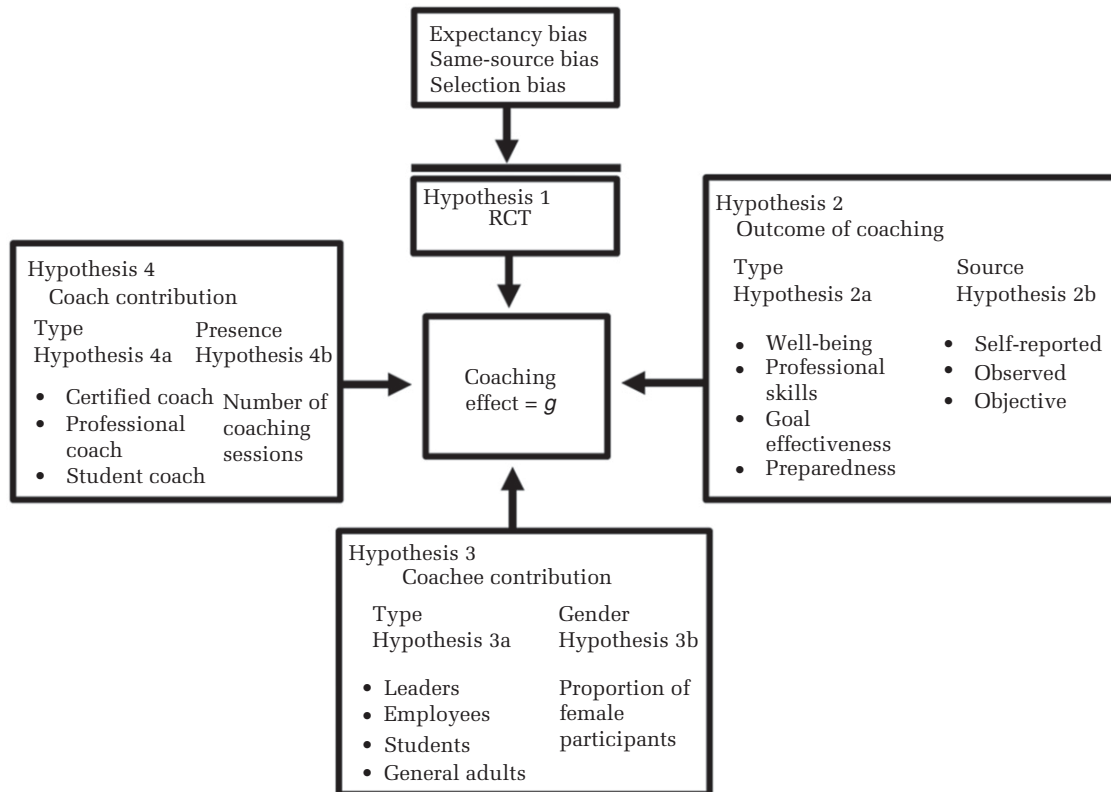
Type of coachee. The question about who benefits (the most) from coaching is important and our meta-analysis can explore for differential effects for different coachees. To the best of our knowledge, direct comparisons between different types of coaching clients, such as students in university settings, general adults, professionals inside organizations, and managers or executives inside organizations, has not been undertaken. From within-sample comparisons, some first indications have been put forward that coachees from a higher educational background, such as university students, benefit more (e.g., Hoven, Ford, Willmot, Hagan, & Siegrist, 2014; Poluka & Kaifi, 2015). Toegel and Nicholson (2005) and Bozer, Baek-Kyoo, & Santora (2015) found a similar small positive effect for same-gender pairs.

We would assume leverage of coaching (from the viewpoint of autonomy as well as trust and safety) is highest when the coachee is relatively separate from an organization (such as in student and general adult populations) and lowest when coaching takes place higher up the hierarchical ranks inside organizations. In view of the male dominance in the business world, in conjunction with gender stereotypes suggesting females are more people oriented, open, and caring (Gray, De Haan, & Bonneywell, 2019), we expect female coachees to be more intrinsically motivated for coaching. Moreover, as the coachee's agenda and motivation is central to coaching and intrinsic, voluntary motivation influences female learners more than male counterparts, both in sports coaching and in traditional education (e.g. Amorose & Horn, 2000; Vecchione, Alessandri, & Marsicano, 2014), we expect a slightly higher effectiveness for female coachees. Thus:

Hypothesis 3a. Workplace coaching will be effective for students, employees, and leaders, and skewed upwards for student and nonleadership coachees.

Hypothesis 3b. Workplace coaching will achieve higher effects for female coachees.

Type of coach and coach presence. Leverage will be highest when the coach is also the coachee's

instructor or line manager, because then the coach's impact tends to stretch beyond the boundaries of formal coaching sessions. For the same reason, we hypothesize external coaches to have lower impact than instructors or managers. In a teaching or laboratory context, leverage should be on an intermediate level: coaches have influence over and authority within the setting, something that external coaches do not have. As far as we know, there are currently very few studies into different types and dosages of coaches. Sue-Chan and Latham (2004) found that students receiving coaching from a faculty instructor achieved significantly higher course grades than students who had been coached by a peer. Students also rated the faculty instructor as more credible than the peer coaches. Liljenstrand and Nebeker (2008) compared 2,231 coaches of different backgrounds and suggested that those with psychology training are more represented in executive coaching and are finding the marketplace more competitive. Finally, Boyce et al., (2010), Grant (2014), and Bozer et al., (2014) have provided some evidence as to the higher impact of a coach with "credibility"—that is, external, qualified, psychologically trained executive coaches. We propose therefore that, compared to those working with student coaches, coachees who receive coaching from an external coach may feel more confident in the credibility and professionalism of the coach, and in the confidentiality of the sessions, which will increase leverage of those coaches and therefore effectiveness.

On first sight, the amount of presence of the coach, expressed by the number of sessions, should increase leverage. However, we have seen strong evidence of coregulation, or "responsive regulation" in the psychotherapy literature (Stiles, Barkham, & Wheeler, 2015), where it is found that therapist and client can coregulate the overall outcome to fit within a good range in terms of number of sessions. In coaching too, coregulation has been found to lead to the result that similar effect sizes are found for different "doses" of coaching or numbers of sessions (see Anthony, Gimbert, & Fultz, 2013; Goff, Goldring, Guthrie, & Bickman, 2014; and, in meta-analysis studies, Sonesh et al., 2015; Theeboom et al., 2014). We postulate:

Hypothesis 4a. Coaching provided by an external, qualified executive coach will demonstrate greater effect sizes than coaching provided by (a) an internal or (b) student coach.

Hypothesis 4b. Coaching impact will be independent of number of coaching sessions provided.
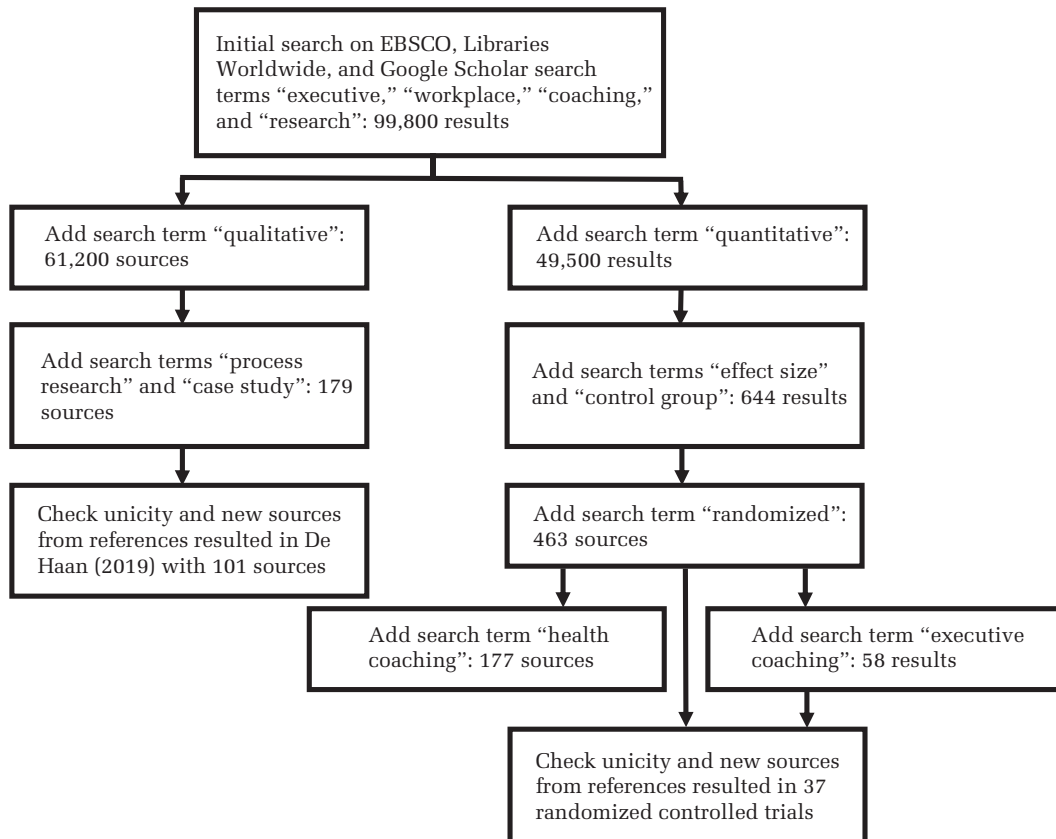
METHODOLOGY

A literature search was carried out using the EBSCO, Google Scholar, and ProQuest databases to identify relevant studies published up to December 31, 2020. The following terms were used sequentially in the search: "executive," "workplace," "coaching," and "research" in combination, then "quantitative," then "effect size" and "control group" in combination, and finally "randomized," with additional terms "health coaching" and "executive coaching" to refine the search (see Figure 2). This search strategy yielded a total of 463 sources, of which 20 were original RCT studies. By cross-checking all references of these sources and the references in earlier systematic reviews and meta-analyses (i.e., Athanasopoulou & Dopson, 2018; Burt & Talati, 2017; Ely, Boyce, Nelson, Zaccaro, Hernez-Broome, & Whyman, 2010; Grover & Furnham, 2016; Jones et al., 2015; Sonesh et al., 2015; Theeboom et al., 2014), we were able to grow the number of original RCT studies in workplace coaching to 37. Details on how many studies were removed at each stage can be found in Figure 2. We checked all remaining articles for our definition of coaching as a series of distinct, one-to-one, client-led, goal-directed conversations, wherein independent (coaching) professionals work with healthy adults. We included one study with adults who had recently been unable to work due to mental disability, because they had been in therapeutic treatment and were being supported by coaches to reintegrate in normal competitive employment (Viering et al., 2015).

Inclusion and Exclusion Criteria

Our selection stays within the definition of workplace or leadership coaching as a series of one-to-one helping conversations, in which the coach is independent and supports a healthy, adult coachee who works on his or her own, mostly work-related goals. For this reason, we have selected all RCTs (with a control group of comparable size to the intervention group) in leadership, life, and workplace coaching, as these have similar coaching methods in common and place the coachee's own challenges and decisions at the center of the process. We have excluded studies in health coaching, sports coaching, and mentoring where the coach is much more in charge, uses a different coaching methodology, and directs the change process by means of instruction or education. These are large domains of practice that would skew any meta-analysis around workplace coaching of independent adults; moreover, we expect a meta-analysis of all these fields to be so diverse that

FIGURE 2
Overview of the Systematic Search and Selection Process

```
┌─────────────────────────────────────┐
│ Initial search on EBSCO, Libraries   │
│ Worldwide, and Google Scholar search │
│ terms "executive," "workplace,"      │
│ "coaching," and "research": 99,800   │
│ results                              │
└─────────────────────────────────────┘
```

Add search term "qualitative": 61,200 sources	Add search term "quantitative": 49,500 results
Add search terms "process research" and "case study": 179 sources	Add search terms "effect size" and "control group": 644 results
Check unicity and new sources from references resulted in De Haan (2019) with 101 sources	Add search term "randomized": 463 sources

Add search term "health coaching": 177 sources	Add search term "executive coaching": 58 results

Check unicity and new sources from references resulted in 37 randomized controlled trials

the PI becomes unwieldy and workplace-related pre-dictions cannot be well calculated.

In summary, we excluded all studies that were not researching workplace or executive coaching in the widest sense. This meant we included work-related, life, career, and executive coaching with adults. We did not exclude any studies on the grounds of the background, interventions, or qualifications of the coach, nor of the nature of the coaching outcomes, nor of the coachees' seniority, gender, or age, lack of realism in the experiment, geography, nor industry. In the overall sample, these dimensions tended to be unique in nearly every study. We selected as many of these aspects as we could as moderating aspects to study in the meta-analysis. We believe our inclusion criteria were straightforward and rigorous: any truly randomized controlled trial (RCT) in adult workplace coaching would be included. See Table 1 for a more detailed description of each study, including the nature of the study, type of coach, type of coachee, types of outcomes, and sample sizes.

We left out more than 100 RCTs that were undertaken in health coaching, because they were all conducted with patients (moreover, they were undertaken in a highly directive, coach-led, educational manner, which would be likely to turn out differently from workplace coaching, where the client's own goals are emphasized). We also left out several mentoring RCTs, which were often nonwork related (e.g., "youth mentoring" or "peer mentoring") and—similar to health coaching—were too advisory and directive to represent coaching interventions. Health coaching and mentoring outcome research have been reviewed in more depth alongside workplace-coaching outcome research in De Haan (2021). Our criteria meant that we did include group coaching (e.g., Green, Oades, & Grant, 2006; Poepsel, 2011; and, partially, De Haan, Gray, & Bonneywell, 2019), life coaching for healthy adults (Green, Grant, & Rynsaardt, 2007; Green et al., 2006; Spence & Grant, 2007) and several studies in which coachees were physically or mentally challenged—for example, by

TABLE 1
Chronologic Overview of All 36 Studies in the Meta-Analysis, Comprising 39 Independent Samples

RCT study (chronological order)	Type of coaching	Coachees	Coaches	N	Outcome measures
Deviney (1994)	Leadership coaching	Middle managers	Higher middle managers	EG1: 14, EG2: 13, CG: 18	360-degree feedback before and after
Taylor (1997)	Coaching in support of education	Graduates	Teaching assistants with coaching training	EG1: 26, EG2: 16, EG3: 26, CG: 16	Medical College Admission Test results, self-scored stress
Grant (2001)	Coaching in support of education	Students	Tutors	Three trials: EG1: 10, CG1: 10; EG2: 9, CG2: 9; EG3: 12, CG3: 12	Several including academic achievement
Miller, Yahne, Moyers, Martinez, and Pirritano (2004)	Coaching in support of education	Health professionals	Motivational interviewing instructors	EG1–4: 24, 33, 34, 26; CG: 23	Work samples and assessment, with 4-, 8-, 12- month follow-up
Green et al. (2006)	Life coaching (in groups)	Healthy adults	Qualified psychologists	EG: 28, CG: 28	Self-scored goal attainment, hope, and well-being, with 3-, 5-, 7- month follow-up
Green et al. (2007)	Workplace coaching	High school students	High school teachers	EG: 25, CG: 24	Self-scored cognitive hardiness, hope, stress
Spence and Grant (2007)	Life coaching	Healthy adults	Qualified psychologists	EG1: 21, EG2: 22, CG: 20	Self-scored goal attainment and well-being
Finn, Mason, and Bradley (2007), based on Finn (2007)	Leadership coaching	Senior managers	Executive coaches	EG: 11, CG: 12	360-degree feedback and self-scored self-efficacy, developmental support, positive affect, openness, and developmental planning
Grant, Curtayne, and Burton (2009)	Leadership coaching	Senior managers	Executive coaches	EG: 21, CG: 20	Self-scored goal attainment, stress, well-being and resilience
Grant, Green, and Rynsaardt (2010)	Workplace coaching	High school teachers	Qualified psychologists	EG: 23, CG: 21	Self-scored goal attainment, stress, well-being, and resilience, with 10-month follow-up
Kochanowski, Seifert, and Yukl (2010)	Leadership coaching	Senior managers	Executive coach	EG: 15, CG: 15	Ratings by subordinates on core influence tactics before and after
Stelter et al. (2011)	Coaching in support of education and career	Students	Qualified coach	EG: 31, CG: 46	Self-scored social recovery and general well-being before and after
Taie (2011)	Workplace coaching	Nurses	Middle managers	EG: 60, CG: 60	Independent rating of life-support skills
Poepsel (2011)	Online life coaching (in groups)	Healthy adults	Qualified coach	EG: 12, CG: 16	Self-scored goal attainment, hope, and well-being
Passmore and Rahman (2012)	Driving-instruction coaching	Healthy adults	Driving instructors	EG: 104, CG: 104	Time in training, first-time pass rate, number of attempts

TABLE 1
(Continued)

RCT study (chronological order)	Type of coaching	Coachees	Coaches	N	Outcome measures
Passmore and Velez (2012)	Driving-instruction coaching	Healthy adults	Driving instructors	EG: 163, CG: 164	Insurance claims, mileage, and driving convictions
Goff et al. (2014)	Leadership coaching	School principals	External school leadership coaches	EG: 26, CG: 26	360-degree feedback before and after
McGonagle, Beatty, and Joffe (2014)	Workplace coaching	General employees	Qualified executive coaches	EG: 23, CG: 25	Self-scored work ability perceptions, job self-efficacy, job satisfaction, burnout, core self-evaluations, and resilience, with 3-month follow-up
Singh, Aggarwal, Tahir, Pucher, and Darzi (2015)	Coaching in support of education	Medical students learning laparoscopic surgical skills	Laparoscopic surgeons/educators	EG: 10, CG: 10	Performance assessment based on two medical operations
Bonrath, Dedy, Gordon, and Grantcharov (2015)	Coaching in support of education	Medical students learning bariatric surgical skills	Bariatric surgeons/educators	EG: 9, CG: 9	Performance assessment based on two videos
Viering et al. (2015)	Reintegration coaching	Adult job seekers with disability	Qualified psychologists	EG: 88, CG: 83	Attainment of competitive employment
Losch, Traut-Mattausch, Mühlberger, and Jonas (2016)	Workplace coaching	Psychology students	Psychology master's students	EG1: 23, EG2: 13, EG3: 27, CG: 21	Self-scored satisfaction, goal attainment, procrastination
Telle et al. (2016)	Workplace coaching	Employees	Job coaches	EG: 51, CG: 34	Self-reported number of sickness absence days, self-scored burnout, life satisfaction, and work attitudes, with 3-month follow-up
Tee, Shearer, and Roderique-Davies (2017)	Workplace coaching	Psychology undergraduates	Peers (undergraduate students)	EG: 19, CG: 11	Self-scored core self-evaluations
Niglio de Figueiredo et al. (2018)	Coaching in support of education	Oncologists being trained in consultation skills	Psychologists	EG: 36, CG: 36	Ratings by experts of consultations with patients before and after comparisons
Allan, Leeson, De Fruyt, and Martin (2018)	Life coaching	Healthy adults	Psychologists	EG: 27, CG: 27	Changes on NEO ("Big 5") personality instrument, with 3-month follow-up
Williams and Lowman (2018)	Leadership coaching	Middle managers	Qualified psychologists	Two trials: EG1: 32, CG1: 32; EG2: 32, CG2: 32	Self-scored leadership competencies and behaviors

TABLE 1
(Continued)

RCT study (chronological order)	Type of coaching	Coachees	Coaches	N	Outcome measures
Alameddine, Englesbe, and Waits (2018)	Coaching in support of education	Medical students learning suturing surgical skills	Surgeons/educators	EG: 8, CG: 8	Faculty and self-scored assessment of suturing based on videos
De Haan et al. (2019)	Leadership coaching	Middle managers	Experienced, qualified executive coaches	EG: 89, CG: 72	360-degree feedback on working alliance and coaching effectiveness before and after, Hogan personality instrument before and after, self-scored well-being, and perceived social support
Dyrbye, Shanafelt, Gill, Satele, and West (2019)	Workplace coaching	Qualified physicians	Experienced, qualified executive coaches	EG: 41, CG: 41	Self-scored distress, well-being, career satisfaction, and meaning in work
McGonagle et al. (2020)	Workplace coaching	Qualified physicians	Experienced, qualified executive coaches	EG: 26, CG: 24	Self-scored burnout, stress, turnover intentions, work engagement, psychological capital, compassion, job self-efficacy, and job satisfaction, with 3-, 6-month follow-up
De Haan et al. (2020)	Workplace coaching	Business school students	Experienced, qualified executive coaches	EG: 76, CG: 95	Self-scored coaching effectiveness, goal attainment, perceived stress
Zanchetta, Junker, Wolf, and Traut-Mattausch (2020)	Workplace coaching	Young employees	Psychology master's students	EG1: 36, EG2: 33, CG: 34	Self-scored goal attainment, imposter phenomenon
Junker, Pömmer, and Traut-Mattausch (2021)	Workplace coaching	Undergraduate students	Psychology master's students	EG: 24, CG: 20	Self-scored goal attainment, stress
Fontes and Dello Russo (2020)	Workplace coaching	Marketing professionals	Experienced, qualified executive coach	EG: 32, CG: 24	360-degree feedback before and after, self-scored psychological capital, job satisfaction, and organizational commitment, with 4-month follow-up
Howlett et al. (2021)	Coaching in support of education	Students	Trained academic coaches	EG: 53, CG: 63	Self-scored metacognitive awareness before and after

Notes: EG = experimental group, CG = control group. If there is more than one EG, then there is always only one group with an added "coaching" condition that has been used for the meta-analysis. More than one CG indicates independent RCTs within the study.

being close to burn out—but where the coaching was still work related and separate from their medical treatment (Duijts, Kant, Van den Brandt, & Swaen, 2008; McGonagle et al., 2014; Viering et al., 2015). The studies undertaken in the related areas of life coaching, driver coaching, and burn-out coaching made use of qualified executive coaches, so they were all broadly using the same methods as the other RCT studies.

The selection process excluded two further studies that started out as rigorous randomized controlled trials and are therefore worth a mention. In each case, the experiment was set up in a realistic setting with executive coachees and proper randomization, but organizational pressures and attrition meant that intervention and control groups could no longer be fully randomized. One of these studies found significant effects and the other had to confirm the null hypothesis. Here is a short summary of each:

- Egan and Song (2005) found that 54 new managers in a multinational retail organization who were coached by external executive coaches received higher line manager ratings than 49 managers in the control group ($d = .6$; $p < .05$). Significant differences on self-scored performance-goal orientation, job satisfaction, and organizational commitment were also found.
- Ungerer et al. (2019) found that 36 new German entrepreneurs in innovative technology-based firms who had received state-sponsored external coaching did not acquire significantly more survival capability than the 57-strong control group, as no significant results were found on any of the indicators of survival capability.

Even though an impressive amount of good research has now been done and many RCTs can be reviewed together, there is still considerable variability remaining between the various studies. Some were undertaken in realistic, corporate settings with executive coachees, while others have student coachees or even student coaches; some have objectively measured or at least intersubjectively validated (observed) outcomes, others have only self-scored outcomes. There are no agreed outcome variables that have been researched throughout the literature (see Table 1). Three key challenges in present-day RCTs are (1) the widespread reliance on *self-score measures*, in about half of the RCTs; (2) the limited statistical power because of *small sample sizes* well under $n = 100$, in four fifths of RTCs; and (3) the focus on leadership—that is, *executive coaching*—in about one fifth of the

RCTs (see also Table 1, which has a summary of all selected studies).

Analysis

Effect sizes. "Effect size" is a quantitative measure to understand the size of the estimated difference between two or more populations. A central tenet of meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009) is that effect sizes can be applied and compared between samples regardless of their methodology and size. In this study, the effect sizes were calculated based on between-group standardized mean differences, which means that original effect sizes from all samples were transformed to a repeated-measures Cohen's delta (d_{ppc2}; Morris, 2008) where means and standard deviations of pre- and post-tests were available. Morris (2008) suggested using a pooled pre-test standard deviation of the experiment and control group to control for effect of the intervention and correcting for small sample sizes; see Figure 3 for the equation used by Morris (2008).

For the calculation of effect sizes, we collected group sizes, all mean scores, and the standard deviations for both groups for at least the time measurements before and after the intervention. There were 11 sources in the sample in which the publication did not include all those data or at least not in a usable format; thus we contacted the authors of these papers and requested the mean scores and standard deviation for the pre- and post-scores of both groups, and we are very grateful for their help. There was one RCT for which, unfortunately, the full data could not be retrieved, and so we had to exclude it from the meta-analysis: Duijts et al., 2008. This left us with 36 of the 37 studies in the meta-analysis, comprising a total of 39 independent samples, because Williams and Lowman (2018) reported two independent samples in their study and Grant (2001) reported three.

FIGURE 3
Definition of Cohen's (1988) Effect Size Extracted from a Pre- and Post-Measurement RCT

$$d_{ppc2} = C_P \left[\frac{(M_{\text{post},T} - M_{\text{pre},T}) - (M_{\text{post},C} - M_{\text{pre},C})}{SD_{\text{pre}}} \right]$$

$$SD_{\text{pre}} = \sqrt{\frac{(n_T - 1)\, SD^2_{\text{pre},T} + (n_C - 1)\, SD^2_{\text{pre},C}}{n_T + n_C - 2}}$$

$$C_P = 1 - \frac{3}{4(n_T + n_C - 2) - 1}$$

Where pre-scores were not available, the reported Cohen's delta were used (in $k = 8$ cases: Finn et al., 2007; Goff et al., 2014; Passmore & Rahman, 2012; Passmore & Velez, 2012; Poepsel, 2011; Singh et al., 2015; Tee et al., 2017; Viering et al., 2015), and, in three other cases, effect sizes had to be transformed from r^2 and r (Stelter et al., 2011; Taie, 2011; Telle et al., 2016). To avoid overestimating the population effect size, given some small sample sizes, we calculated Hedges' *g*, which has the same small-sample correction as d_{ppc2} (Borenstein et al., 2009).

We conducted a power analysis for random-effects models to identify the number of studies required to achieve a power of $> 80\%$ that is recommended to identify a true effect (Ellis, 2010). We used a conservative effect size of $d = .20$ to allow studies with low effects to be identified, and the sample size was set at $n = 56$, which was the average number of participants in the meta-analysis by Theeboom et al. (2014) when outliers had been controlled for. This power-analysis revealed that $k = 27$ would be sufficient to assure power over 80%.

Model and heterogeneity. The meta-analysis in this paper is conducted on RCTs of different participants, methodologies, and outcomes. A random effects model in meta-analysis is appropriate to use when the effect sizes are assumed to fluctuate between studies due to uncontrolled variables, such as differences in the methodology of the interventions, differences in participants, etc. (Borenstein 2019). Given this heterogeneity in intervention content, we decided in advance to pool data using a random-effects model, with all effect sizes weighted by the reciprocal of the sampling variances, as suggested by Borenstein, Hedges, Higgins, & Rothstein (2010). Furthermore, the Hartung–Knapp–Sidik–Jonkman (HKSJ) adjustment was used to correct the random-effects model and create more robust estimates (Hartung & Knapp, 2001; Sidik & Jonkman, 2002). The HKSJ adjustment often results in lower error rates in a meta-analysis that contains a smaller number of studies and studies with similar sample sizes (IntHout, Ioannidis, & Borm, 2014). The analysis of the moderating relationships (Hypotheses 2–4) is conducted on smaller subsets of studies ($k4$–$k27$) and the HKSJ adjustment is therefore a suitable and preferable adjustment model for the analyses in this study.

As recommended by Borenstein et al. (2009), heterogeneity between samples was measured by calculating both the classical Q statistic and the I^2 statistic. The most used heterogeneity statistic is Q, which calculates a weighted sum of the square distances of the observed effects from the null hypothesis of equality of the effects. The Q statistic provided a significance test for between-sample heterogeneity, whereas the value for I^2 represented the percentage of between-sample variance in effect sizes that can be attributed to between-sample heterogeneity rather than within-sample variability (Borenstein et al., 2009). The software used for the analysis was MIX 2.0, developed by Leon Bax (2016).

Inclusion variables and outcome categorization. As Table 1 shows, most RCTs contained multiple outcome variables. Thus, a four-step selection process was implemented to categorize outcomes and select the variables to include in the main meta-analysis. The first step was to assess if a variable was characterized as the main outcome variable by the authors. If, instead, a sample had multiple variables in focus, the second step consisted of selecting an objective, or (next) an observed variable, such as ratings by direct reports or peers, instead of a self-reported variable, to reduce same-source biases. If this did not yield a clear outcome variable for that sample, in a third step, we undertook an analysis of the variables in the wider literature to assess which were more reliable than others. If none of these three steps yielded a single remaining outcome measure, we randomly selected one and only one of the remaining variables, to prevent violation of the independent sample assumption (Borenstein et al., 2009).

When we later tested explicitly for differences between self-reported, observed, and objective outcome variables, and a single sample provided variables in more than one of these categories, then they were all selected for this test, so as to include the widest possible range of samples.

Moderator variables. Moderator variables were investigated using subgroup analyses for sets of samples that differed in terms of the samples of participants and outcome variables. Moderating variables are considered to be present when the value for I^2 is close to or above 50% (Higgins & Thompson, 2002). Subgroup analysis and multiple meta-regressions were performed using the R packages Meta (Schwarzer, 2007) and Metafor (Viechtbauer, 2010). The subgroup analysis pooled the effect size of each subgroup and compared the effect sizes of each subgroup with each other using a random effects model (Borenstein & Higgins, 2013). Our multiple meta-regression models used proportion of females and number of sessions as predictors of the effect size (this could be done for the $k = 33$ samples that provided these data). These multiple-regression models were controlled by HKSJ adjustments and only performed on subgroups with $k \geq 10$, as recommended by Borenstein et al. (2009: 188).

RESULTS

Data Check and Meta-Analysis

The Results section is outlined after the four hypotheses, and will present the impact of coaching on types of measurements of outcomes, nature of outcomes, coachees, and coaches. The random effects meta-analysis of executive coaching against control groups yielded an overall weighted effect size for all outcomes that was in the moderate range, reflecting an advantage of coaching over control groups, with Hedges' $g = .59$ [0.45, 0.74], $t = 8.42$, $p < .001$ (95% PI [-0.19 to 1.38]; Figure 4). All results of the meta-analyses are summarized in Table 2.

An exclusion-sensitivity assessment of the different studies, with the aim of identifying how specific studies impacted the overall results, indicated that the studies with the most participants (Passmore & Rahman, 2012; Passmore & Velez, 2012) skewed the data significantly. An analysis of the whole sample with the two articles removed showed a higher effect size, $g = .63$ [0.48, 0.77], $t = 8.69$, $p < .001$. This improves the prediction interval, which is now PI [-0.15 to 1.40]. The two studies excluded from further analysis form part of a group of four studies in which the coaches were also instructors or line managers of the coachees, and where it can be argued that coaches provide a lot more than just "coaching sessions" (Deviney, 1994; Passmore & Rahman, 2012; Passmore & Velez, 2012; Taie, 2011). We tested what would happen to the overall effect size if these studies were taken out. Excluding any combination of these studies did not have a large impact on the effect size, with the lowest effect size found when they are all taken out, $g = .59$ [0.46, 0.73], $t = 8.93$, $p < .001$, and PI [-0.07 to 1.26]. Taking any of these samples out did improve the PI, which confirms that they were probably outliers and they have limited relevance to workplace and executive coaching. Passmore and Rahman (2012) and Passmore and Velez (2012) were removed from further subgroup analysis where the overall weight is impacted more, because the coaches were primarily instructors.

In sum, Hypothesis 1 is clearly supported, although the 95% PI only nearly excludes zero, especially when studies with coaches who are also managers or instructors are taken out.

Types of Measurements of Outcomes

While most studies were focusing on well-being dependent variables ($k = 15$), there were also numerous studies with a focus on professional skills ($k = 13$) or goal attainment ($k = 10$). A few studies did also focus on measuring the impact of coaching on general preparedness, such as self-efficacy or psychological capital ($k = 9$). All types of measures were significantly impacted by coaching (see Table 2). A subgroup analysis did not indicate that the effect of coaching significantly differed between any of the variables. This supports Hypothesis 2a regarding coaching outcomes.

Self-reported measures were the most reported measure ($k = 27$), followed by objective ($k = 8$) and observed measures ($k = 8$), and, as can be seen in Table 2, all three types of measures were significantly impacted by coaching. Although the objective outcomes had a high pooled g, they did not significantly differ from either self-reported outcomes ($z = -0.45$ [-0.51, 0.81], $p = .65$) or observed outcomes ($z = -1.51$ [-1.14; -0.00], $p = .049$) after the confidence interval being corrected by the Bonferroni correction (0.05 / 3 = 0.017). However, coaching does have a significantly higher effect on self-reported measures compared to observed measures ($z = 2.63$, [0.10, 0.71], $p = .008$). This supports Hypothesis 2b regarding self-scored variables.
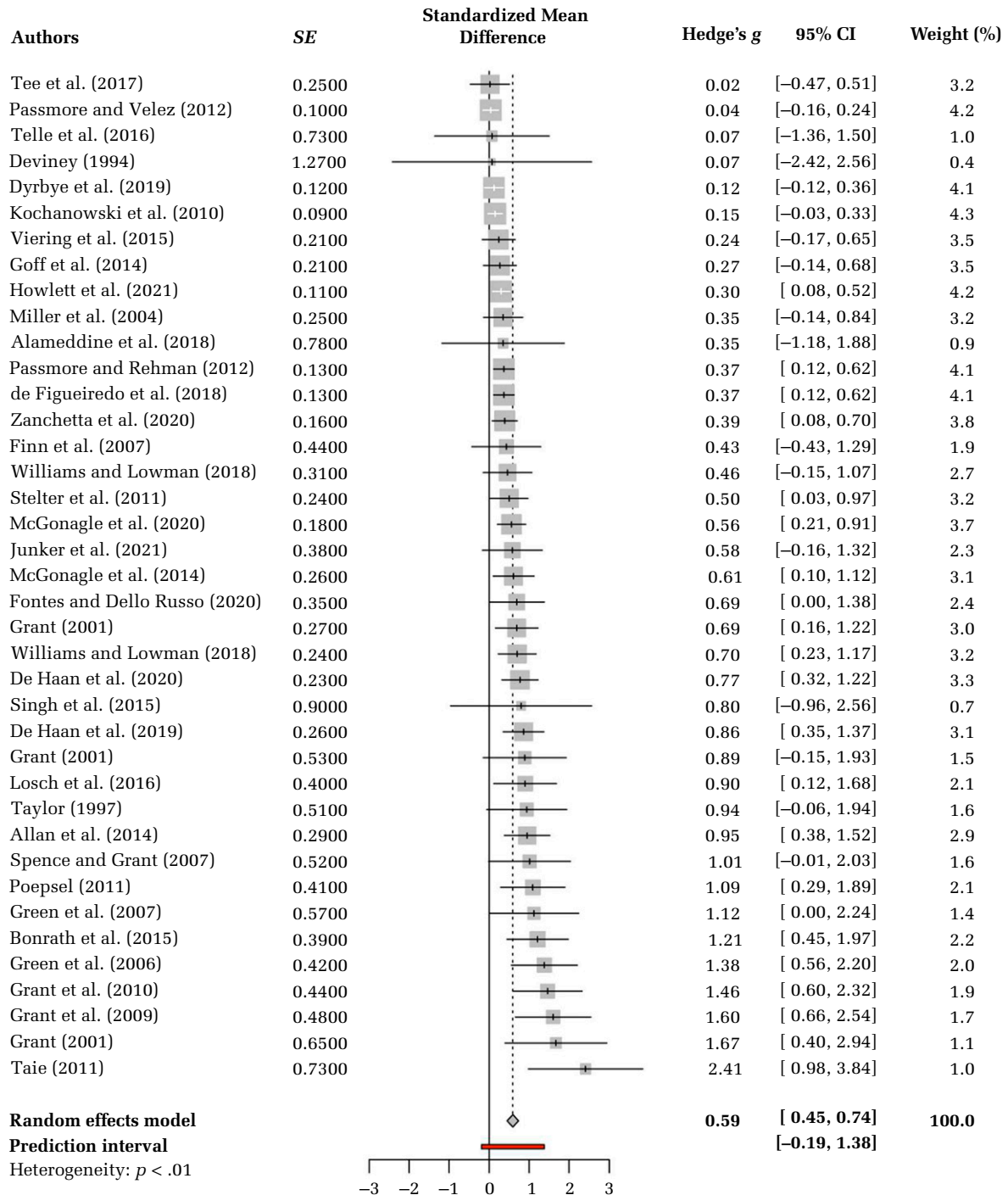
Coachees

Coaching studies were mostly conducted on employees ($k = 11$) and students ($k = 13$), followed by leaders ($k = 7$), and, lastly, a more general adult population ($k = 6$). Coaching had a significant impact on all participant groups; however, the significant effect disappeared on leadership coaching when a Bonferroni correction was implemented (0.05 / 4 = 0.013). Furthermore, a subgroup analysis did not identify significant effect differences between any of the participant subgroups. This lends support to Hypothesis 3a, which stated that we expect lower overall effects on leaders and managers.

Coaches

The majority of the studies included in this meta-analysis used coaches with some level of certificate or accreditation ($k = 15$). Other studies had trained internal coaches ($k = 8$) or students as coaches ($k = 4$). From the analysis, we learn that both certified coaches and internal coaches had a significant effect, while the effect sizes of student coaches was not significantly different from zero (see Table 2). However, a subgroup analysis could not identify any significant differences in the effect between the types of coaches. This means

FIGURE 4
Forest Plot Showing the Effect of RCT on Coaching Outcomes for Increasing Effect Size

Authors	*SE*	Standardized Mean Difference	Hedge's *g*	95% CI	Weight (%)
Tee et al. (2017)	0.2500		0.02	[−0.47, 0.51]	3.2
Passmore and Velez (2012)	0.1000		0.04	[−0.16, 0.24]	4.2
Telle et al. (2016)	0.7300		0.07	[−1.36, 1.50]	1.0
Deviney (1994)	1.2700		0.07	[−2.42, 2.56]	0.4
Dyrbye et al. (2019)	0.1200		0.12	[−0.12, 0.36]	4.1
Kochanowski et al. (2010)	0.0900		0.15	[−0.03, 0.33]	4.3
Viering et al. (2015)	0.2100		0.24	[−0.17, 0.65]	3.5
Goff et al. (2014)	0.2100		0.27	[−0.14, 0.68]	3.5
Howlett et al. (2021)	0.1100		0.30	[0.08, 0.52]	4.2
Miller et al. (2004)	0.2500		0.35	[−0.14, 0.84]	3.2
Alameddine et al. (2018)	0.7800		0.35	[−1.18, 1.88]	0.9
Passmore and Rehman (2012)	0.1300		0.37	[0.12, 0.62]	4.1
de Figueiredo et al. (2018)	0.1300		0.37	[0.12, 0.62]	4.1
Zanchetta et al. (2020)	0.1600		0.39	[0.08, 0.70]	3.8
Finn et al. (2007)	0.4400		0.43	[−0.43, 1.29]	1.9
Williams and Lowman (2018)	0.3100		0.46	[−0.15, 1.07]	2.7
Stelter et al. (2011)	0.2400		0.50	[0.03, 0.97]	3.2
McGonagle et al. (2020)	0.1800		0.56	[0.21, 0.91]	3.7
Junker et al. (2021)	0.3800		0.58	[−0.16, 1.32]	2.3
McGonagle et al. (2014)	0.2600		0.61	[0.10, 1.12]	3.1
Fontes and Dello Russo (2020)	0.3500		0.69	[0.00, 1.38]	2.4
Grant (2001)	0.2700		0.69	[0.16, 1.22]	3.0
Williams and Lowman (2018)	0.2400		0.70	[0.23, 1.17]	3.2
De Haan et al. (2020)	0.2300		0.77	[0.32, 1.22]	3.3
Singh et al. (2015)	0.9000		0.80	[−0.96, 2.56]	0.7
De Haan et al. (2019)	0.2600		0.86	[0.35, 1.37]	3.1
Grant (2001)	0.5300		0.89	[−0.15, 1.93]	1.5
Losch et al. (2016)	0.4000		0.90	[0.12, 1.68]	2.1
Taylor (1997)	0.5100		0.94	[−0.06, 1.94]	1.6
Allan et al. (2014)	0.2900		0.95	[0.38, 1.52]	2.9
Spence and Grant (2007)	0.5200		1.01	[−0.01, 2.03]	1.6
Poepsel (2011)	0.4100		1.09	[0.29, 1.89]	2.1
Green et al. (2007)	0.5700		1.12	[0.00, 2.24]	1.4
Bonrath et al. (2015)	0.3900		1.21	[0.45, 1.97]	2.2
Green et al. (2006)	0.4200		1.38	[0.56, 2.20]	2.0
Grant et al. (2010)	0.4400		1.46	[0.60, 2.32]	1.9
Grant et al. (2009)	0.4800		1.60	[0.66, 2.54]	1.7
Grant (2001)	0.6500		1.67	[0.40, 2.94]	1.1
Taie (2011)	0.7300		2.41	[0.98, 3.84]	1.0
Random effects model			**0.59**	**[0.45, 0.74]**	**100.0**
Prediction interval				**[−0.19, 1.38]**	
Heterogeneity: *p* < .01					

−3 −2 −1 0 1 2 3

The figure shows the standard error of the effect size (*SE*), the standardized mean difference (Hedges' *g*), the 95% confidence interval (CI), and the weight of each study to the overall effect. Each study of the main analysis is displayed and sorted on their standardized mean difference. The overall effect and the PI are shown at the bottom.

TABLE 2
Results of the Principal Meta-Analyses

	Descriptive			Random effect			Heterogeneity		
	k	*n*	*g*	Lower	Upper	*p*	Q	*I²* (%)	*T²*
Full sample	39	2,528	.59	0.45	0.74	< .001	87.26***	55.5	0.14 [0.02, 0.47]
Source of outcome									
Self-reported outcomes	27	1,516	.62	0.46	0.78	< .001	61.02***	57.4	0.10 [0.01, 0.20]
Observed outcomes	8	387	.29	0.15	0.43	.002	4.25	0.0	0.01 [0.00, 0.05]
Objective outcomes	8	331	.89	0.25	1.53	.015	15.65*	61.7	0.37 [0.01, 2.55]
Nature of outcome									
Well-being	15	788	.61	0.37	0.86	< .001	61.90***	77.4	0.15 [0.02, 0.44]
Professional skills/performance	13	533	.67	0.33	1.02	.001	27.66**	56.6	0.24 [0.01, 0.81]
Goal effectiveness	10	563	.82	0.46	1.19	< .001	21.64*	58.4	0.16 [0.01, 0.70]
Preparedness	9	510	.42	0.15	0.70	.007	15.61*	49	0.11 [0.00, 0.70]
Participants									
Leaders	7	343	.54	−0.01	1.09	.049	14.76*	66.1	0.19 [0.01, 1.71]
Employees	11	580	.67	0.32	1.02	.002	26.64**	62.5	0.24 [0.02, 0.99]
Students	13	716	.56	0.34	0.78	< .001	14.98	19.9	0.08 [0.00, 0.27]
General adults	5	245	.84	0.26	1.42	.015	9.42	57.5	0.11 [0.00, 1.38]
Type of coach									
Certified coaches	15	1,011	.62	0.40	0.83	< .001	31.16**	55.1	0.10 [0.01, 0.31]
Professionally trained coaches	8	373	.88	0.32	1.45	.008	13.07	46.4	0.27 [0.00, 1.70]
Student-coaches	4	158	.51	−0.06	1.09	.062	1.49	0.0	0.2 [0.00, 2.52]

Notes: Table shows number of samples *k*, total sample size *n*, effect sizes Hedges' *g*, 95% confidence intervals for *g*; and heterogeneity statistics in terms of Q, *I²*, and Tau squared (*T²*), with its 95% confidence intervals.

that Hypothesis 4a is not supported, yet there is some support for Hypothesis 4b.

Multiple Meta-Regression

Multiple meta-regressions were conducted to understand the impact of the number of coaching sessions and the proportion of female clients in the research sample. The multiple meta-regression indicated that the predictors significantly explained 61.99% of the effect size heterogeneity ($r^2 = 61.99$), $f(2, 29) = 5.71$, $p < .01$, where the proportion of females had a significant unique prediction of the effect ($b = 0.007$, $SE = 0.003$, $p < .01$). This signifies that a 10% increase in female proportion is associated with an effect size increase of 0.07. However, the number of coaching sessions did not have a unique prediction of the overall effect ($b = 0.04$, $SE = 0.02$, $p > .05$). The number of coaching sessions ranged from 1 to 16, with an average of seven sessions. A regression on effect size confirms that effect sizes are stable between four to eight interventions. The studies with higher effect sizes with interventions between four to eight are studies with fewer participants. There seem to be a few studies around 10 interventions with a higher effect size, but these studies have a lower weight of the meta-analysis.

The number of coaching sessions did, however, significantly predict the effect size for self-reported measures only. In fact, the full regression showed $r^2 = 70.54$, $f(2, 23) = 4.38$, $p < .01$, with number of coaching sessions ($b = 0.05$, $SE = 0.03$, $p < .05$); and, again, the proportion of females uniquely predicted this effect size as well ($b = 0.01$, $SE = 0.003$, $p < .05$).

Hypotheses 3b and 4b were both supported, with females benefiting slightly but significantly more from coaching in these samples, and numbers of sessions having nearly no impact on effectiveness.

Publication Bias

We performed an analysis of potential publication bias by conducting a funnel plot and Egger's regressions analysis (Duval & Tweedie, 2000). The funnel plot in Figure 5 suggests that there is funnel asymmetry among the studies. Egger's regression test further supported that there is an asymmetry, $b = 1.99$, $t(37) = 6.57$, $p < .001$. We corrected the meta-analysis using the trim-and-fill method (Duval & Tweedie, 2000), due to the asymmetry of the studies. The trim-and-fill method identified 16 studies that may have inflated the effect size, and the trim-and-fill-corrected analysis revealed a lower overall effect size, $g = .32$ [0.13, 0.52], $t = 3.35$, $p < .01$ [PI = −0.99 to 1.63]. It is therefore likely that the overall reported effect size ($g = .59$) is somewhat overestimated due to

FIGURE 5
Contour-Enhanced Funnel Plot (Coaching Effect)
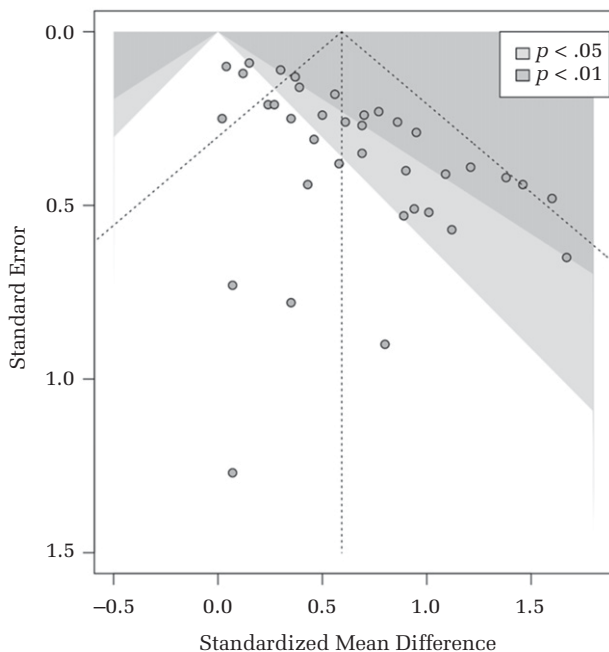
The funnel plot illustrates the standardized mean difference using Hedges' *g* by the standard error of the effect size. The shaded areas show studies with significant effect sizes (*p* < .01 and *p* < .05, separately). Authors of the studies near the *p* = .05 threshold might have calculated the effect size differently leading to significant results. One indicator of publication bias is a negative relationship between standard error and effect size (Kepes, Banks, McDaniel, & Whetzel, 2012), which appears to be the case and supports the indication of publication bias.

publication biases, like the presence of publication bias in psychotherapy (Cuijpers et al., 2020).

DISCUSSION

We believe that this meta-analysis can give us confidence regarding the effectiveness of coaching programs. We have found convincing indications for moderate effectiveness, with best estimate $g = .59$ and 95% PI [-0.07 to 1.26] (taking the four articles that report on coaching by managers or instructors out of the sample). This estimate is just below what is usually found for psychotherapy, with latest estimates of *g* being around .73 (see, e.g., Cuijpers et al., 2020) while our PI is narrower than the ones found in psychotherapy. A lower effect size for coaching can be expected from the fact that psychotherapy has higher problem pressure, tends to have more structure around the sessions with more leverage (power) residing in psychotherapists, and a higher

frequency and therefore intensity. Given that an effect size of 0.5 is large enough to be seen by the naked eye, such as the difference in height between 14- and 18-year-old children (Cohen, 1988), this is a sizeable effect indeed.

It is important to realize that there is a high variability in the sample and that most of the demonstrated effectiveness relies on self-reported outcomes by the coachees, which are notoriously biased toward false positives (Grover & Furnham, 2016). Nevertheless, when we tested other outcome measures, we still found significantly positive effect sizes (see Table 2). Thus, when understanding that the effect of coaching can easily be seen by the "naked eye," it can be pointed out that this is not just the eye of the coachees themselves. Coaching outcomes are still significant but somewhat less visible to direct reports, peers, and line managers ("observed variables").

This meta-analysis has not just shown a clear PI for effectiveness of workplace coaching based on RCT studies, it has also identified a number of significant moderators of this effectiveness, in terms of types of coaches, coachees, and outcomes (see Figure 1). Finally, we confirm the helpfulness of our coregulation model, which is, at the same time, gaining traction in process studies in the helping professions, which can fill in further gaps, studying moment-by-moment changes during the sessions (e.g., Ianiro et al., 2015, and Erdös & Ramseyer, 2021). We will now look at the conceptual and practical inferences that we believe can be made on the basis of the meta-analysis.

Methodological and Conceptual Contributions

Making use of a coregulation model of coaching effectiveness (Erdös & Ramseyer, 2021; Watson & Wiseman, 2021) has helped to bring clarity to our findings. We hypothesized that coregulation based on mutual leverage, trust, and motivation between coach and coachee would be the key predictor of coaching effectiveness. This model successfully predicted five disparate significant relationships in this study (Hypotheses 2b − 4b), and is therefore a promising perspective on the effectiveness of coaching relationships. Coregulation may also explain our finding that the population effect size *g* in coaching RCTs is lower than what has been found recently in psychotherapy (Cuijpers et al., 2020), because therapy does enjoy more leverage in terms of therapist credibility, frequency of sessions, and problem pressure.

We find positive effect sizes for many different measures of outcome, which confirms the idea that coaching is an effective intervention on a wide

variety of impacts. This matches the fact that, in practice, coaching works with highly tailored (contracted) coaching objectives, ranging from personal well-being and work efficiencies through to organizational outputs. We believe that our high effect sizes for objective outcome variables (namely, $g = .89$; see Table 2) were skewed by a single study by Taie (2011), which had a large sample and made use of internal coaches who were also line managers. In any case, the difference between objective, self-scored, and observed outcomes was not significant.

Directions for Future Research

We think it is possible to make further predictions with the coregulation model and to further strengthen it by looking at how it predicts a variety of settings, impacts, and results of coaching. It seems that, across a wide variety of geographies, industries, and professions, coach and coachee will manage to adapt to the circumstances of the work, the objectives they agree, and the permitted length of the assignment to optimize their interactions. In this process, coach and coachee implicitly respond and subtly negotiate "good enough" terms for outcomes, optimizing outcomes even in the face of limitations such as a low number of sessions or lack of coaching or leadership experience, such as with student coachees and student coaches.

The PI is clearly influenced by the nature of the coaching and becomes narrower if coaching by line managers is taken out. We would suggest that future meta-analysis studies exclude coaching by managers and that they study more dimensions of coaching by qualified coaches and student coaches, such as moderation by the nature of training or approach of the coaches; age and years-of-experience of the coaches; the quality of the relationship between coachee and coach; and methodological variables, such as quality of randomization.

In future meta-analytic research, we believe it is important to test the coregulation model and to base further and perhaps more subtle predictions on it—for example, the differential impact of coaching with *managerial level* of the coachee, or with *level of credibility* of the coach. Some quasi-experimental studies have already been done in these areas and they support our model (see, e.g., Agarwal, Angst, & Magni, 2009, for managerial level, and Bozer et al., 2014, for coach credibility), but there are no RCTs as far as we can find. We also need more studies that link leverage to intrinsic motivation and empowerment of the coachee, which is a factor that may go a long way in explaining our differential results for male and female coachees. Another aspect worth testing is that our findings and model seem to support the hypothesis that more challenging coaches are effective for senior leaders, as they can themselves be challenging and powerful, and may therefore need more leverage for the coaching relationship to work well.

Contributions to Practice and Coach Development

This meta-analysis has brought significant support for the effectiveness of coaching as a tailored form of personal and organizational development for leaders and others in organizations. It supports a long-standing debate that management learning and organizational development should be more tailor-made, customized, and flexible in the current VUCA world (LeBlanc, 2018).

We found some confirmations of unique findings in the literature, such as that the number of sessions does not matter much (see also Anthony et al., 2013). The findings of the meta-regressions suggest that the gender distribution of the samples significantly impacts the effect, although this may be skewed by one large study that had only female coachees (De Haan et al., 2019). The effect of gender was small but significant. It is worth studying more, particularly when considered from the emancipatory and intrinsically motivational aspects of leadership coaching. Coaching studies undertaken with a single gender may be skewed, so we need more within-sample comparisons.

Regarding the noneffect of numbers of sessions, it seems that coachees and coaches coregulate to maximize their take-up of coaching for at least the range of sessions between four and eight, so that they adjust the total number of sessions to what they need, or else make sure that they achieve what they can in the number of sessions provided, no matter the precise number of sessions, over a wide range (see Stiles et al., 2015, for an extensive discussion of this coregulation phenomenon in psychotherapy). The earlier meta-analyses by Theeboom et al. (2014) and Sonesh et al. (2015) also found that effect sizes were hardly influenced by the number of sessions. This means we should be careful not to think about the number of sessions as the "dosage" of coaching; for example, even in shorter contracts with fewer sessions, a high effect size can be reached. Our finding also does not preclude higher effects in individual contracts when they are extended to more sessions (something that has indeed been found; see De Haan, Molyn, & Nilsson, 2020): coachees can adjust again to take something

more out of every additional session. In this regard, it may be relevant that we find that the number of sessions is a significant moderator when measured purely on self-reported variables.

A strong message comes out of this meta-analysis to practicing coaches and developers of coaches. It is that working on the alliance, trust, and mutual influence pays off in terms of coaching effectiveness. So, anything that can be done to strengthen the cocreated relationship should be worthwhile, such as careful contracting and working flexibly with a variety of requested outcomes, agreeing on client-initiated goals and tasks (Gessnitzer & Kauffeld, 2015), and providing however many sessions the client can agree to. Provided outcomes and session numbers are carefully contracted, there seems to be a "sweet spot" in the range between four and eight sessions wherein coaching programs could make savings with limited detrimental results in terms of effectiveness. We also believe that coach behaviors such as dominant-friendliness (Ianiro et al., 2013) and perceived empathy can further strengthen this alliance. All these aspects can be the focus of coach training and development as well as supervision.

REFERENCES

Agarwal, R., Angst, C. M., & Magni, M. 2009. The performance effects of coaching: A multilevel analysis using hierarchical linear modeling. *International Journal of Human Resource Management*, 20: 2110–2134.

Alameddine, M. B., Englesbe, M. J., & Waits, S. A. 2018. A video-based coaching intervention to improve surgical skill in fourth-year medical students. *Journal of Surgical Education*, 75: 1475–1479.

Allan, J., Leeson, P., De Fruyt, F., & Martin, S. 2018. Application of a 10-week coaching program designed to facilitate volitional personality change: Overall effects on personality and the impact of targeting. *International Journal of Evidence Based Coaching and Mentoring*, 16: 80–94.

Amorose, A. J., & Horn, T. S. 2000. Intrinsic motivation: Relationships with collegiate athletes' gender, scholarship status, and perceptions of their coaches' behavior. *Journal of Sport & Exercise Psychology*, 22: 63–84.

Anthony, A., Gimbert, B., & Fultz, D. 2013. The effect of e-coaching attendance on alternatively certified teachers' sense of self-efficacy. *Journal of Technology and Teacher Education*, 21: 277–299.

Athanasopoulou, A., & Dopson, S. 2018. A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most? *Leadership Quarterly*, 29: 70–88.

Bax, L. 2016. *MIX 2.0: Professional software for meta-analysis in Excel* (Version 2.0.1.5). Retrieved from https://www.meta-analysis-made-easy.com

Bonrath, E. M., Dedy, N. J., Gordon, L. E., & Grantcharov, T. P. 2015. Comprehensive surgical coaching enhances surgical skill in the operating room. *Annals of Surgery*, 262: 205–212.

Borenstein, M. 2019. Heterogeneity in meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis*, (3rd edn): 453–470. New York, NY: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. 2009. *Introduction to meta-analysis*. New York, NY: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1: 97–111.

Borenstein, M., & Higgins, J. P. 2013. Meta-analysis and subgroups. *Prevention Science*, 14: 134–143.

Boyce, L. A., Jackson, R. J., & Neal, L. J. 2010. Building successful leadership coaching relationships: Examining impact of matching criteria in a leadership coaching program. *Journal of Management Development*, 29: 914–931.

Bozer, G., Baek-Kyoo, J., & Santora, J. C. 2015. Executive coaching: Does coach-coachee matching based on similarity really matter? *Consulting Psychology Journal*, 67: 218–233.

Bozer, G., Sarros, J. C., & Santora, J. C. 2013. The role of coachee characteristics in executive coaching for effective sustainability. *Journal of Management Development*, 32: 277–294.

Bozer, G., Sarros, J. C., & Santora, J. C. 2014. Academic background and credibility in executive coaching effectiveness. *Personnel Review*, 43: 881–897.

Burt, D., & Talati, Z. 2017. The unsolved value of executive coaching: A meta-analysis of outcomes using randomised control trial studies. *International Journal of Evidence Based Coaching and Mentoring*, 15: 17–24.

Cohen, J. 1988. *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Erlbaum.

Cuijpers, P., Karyotaki, E., De Wit, L., & Ebert, D. D. 2020. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. *Psychotherapy Research*, 30: 279–293.

De Haan, E. 2008. *Relational coaching: Journeys towards mastering one-to-one learning*. Chichester, U.K.: Wiley.

De Haan, E. 2021. *What works in executive coaching: Understanding outcomes through quantitative*

research and practice-based evidence. New York, NY: Routledge.

De Haan, E., Grant, A., Burger, Y., & Eriksson, P.-O. 2016. A large-scale study of executive coaching outcome: The relative contributions of working relationship, personality match, and self-efficacy. *Consulting Psychology Journal*, 68: 189–207.

De Haan, E., Gray, D. E., & Bonneywell, S. 2019. Executive coaching outcome research in a field setting: A near-randomized controlled trial study in a global healthcare corporation. *Academy of Management Learning & Education*, 18: 1–25.

De Haan, E., Molyn, J., & Nilsson, V. O. 2020. New findings on the effectiveness of the coaching relationship: Time to think differently about active ingredients? *Consulting Psychology Journal*, 72: 155–167.

Deviney, D. E. 1994. *The effect of coaching using multiple rater feedback to change supervisor behavior* (Doctoral dissertation, Nova University). Available from ProQuest dissertations and theses database. (UMI No. 9413262)

Duijts, S. F. A., Kant, I., Van den Brandt, P. A., & Swaen, G. M. H. 2008. Effectiveness of a preventive coaching intervention for employees at risk for sickness absence due to psychosocial health complaints: Results of a randomized controlled trial. *Journal of Occupational and Environmental Medicine*, 50: 765–776.

Duval, S., & Tweedie, R. 2000. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56: 455–463.

Dyrbye, L. N., Shanafelt, T. D., Gill, P. R., Satele, D. V., & West, C. P. 2019. Effect of a professional coaching intervention on the well-being and distress of physicians: A pilot randomized clinical trial. *JAMA Internal Medicine*, 179: 1406–1414.

Egan, T., & Song, Z. 2005. *A longitudinal quasi-experiment on the impact of executive coaching*. Paper presented at the 20th annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Ellis, P. D. 2010. *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, U.K.: Cambridge University Press.

Ely, K., Boyce, L. A., Nelson, J. K., Zaccaro, S. J., Hernez-Broome, G., & Whyman, W. 2010. Evaluating leadership coaching: A review and integrated framework. *Leadership Quarterly*, 21: 585–599.

Erdös, T., & Ramseyer, F. T. 2021. Change process in coaching: Interplay of nonverbal synchrony, working alliance, self-regulation, and goal attainment. *Frontiers in Psychology*, 12. doi: 10.3389/fpsyg.2021.580351

Finn, F. A. 2007. *Leadership development through executive coaching: The effects on leaders' psychological states and transformational leadership behaviour* [Unpublished doctoral dissertation]. Queensland University of Technology, Brisbane.

Finn, F. A., Mason, C. M., & Bradley, L. M. 2007. *Doing well with executive coaching: Psychological and behavioural impacts*. Paper presented at the Academy of Management annual meeting proceedings, Philadelphia, PA.

Fontes, A., & Dello Russo, S. 2020. An experimental field study on the effects of coaching: The mediating role of psychological capital. *Applied Psychology*, 70: 459–488.

Gessnitzer, S., & Kauffeld, S. 2015. The working alliance in coaching: Why behaviour is the key to success. *Journal of Applied Behavioral Science*, 51: 177–197.

Goff, P., Goldring, E., Guthrie, J., & Bickman, L. 2014. Changing principals' leadership through feedback and coaching. *Journal of Educational Administration*, 52: 682–704.

Grant, A. M. 2001. *Towards a psychology of coaching: The impact of coaching on metacognition, mental health and goal attainment* (Doctoral dissertation, Macquarie University). Retrieved from https://files.eric.ed.gov/fulltext/ED478147.pdf

Grant, A. M. 2012. An integrated model of goal-focused coaching: An evidence-based framework for teaching and practice. *International Coaching Psychology Review*, 7: 146–165.

Grant, A. M. 2014. Autonomy support, relationship satisfaction and goal focus in the coach–coachee relationship: which best predicts coaching success? *Coaching*, 7: 18–38.

Grant, A. M., Curtayne, L., & Burton, G. 2009. Executive coaching enhances goal attainment, resilience and workplace well-being: A randomized controlled study. *Journal of Positive Psychology*, 4: 396–407.

Grant, A. M., Green, L. S., & Rynsaardt, J. 2010. Developmental coaching for high school teachers: Executive coaching goes to school. *Consulting Psychology Journal*, 62: 151–168.

Gray, D. E., De Haan, E., & Bonneywell, S. 2019. Coaching the "ideal worker": Female leaders and the gendered self in a global corporation. *European Journal of Training and Development*, 43: 661–681.

Green, L. S., Grant, A. M., & Rynsaardt, J. 2007. Evidence-based life coaching for senior high school students: Building hardiness and hope. *International Coaching Psychology Review*, 2: 24–32.

Green, L. S., Oades, L. G., & Grant, A. M. 2006. Cognitive-behavioural, solution-focused life coaching: Enhancing goal striving, well-being, and hope. *Journal of Positive Psychology*, 1: 142–149.

Grover, S., & Furnham, A. 2016. Coaching as a developmental intervention in organizations: A systematic review of its effectiveness and the mechanisms underlying it. *PLoS One*, 11. doi: 10.1371/journal.pone.0159137

Hartung, J., & Knapp, G. 2001. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, 20: 3875–3889.

Higgins, J., & Thompson, S. G. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21: 1539–1558.

Hoven, H., Ford, R., Willmot, A., Hagan, S., & Siegrist, J. 2014. Job coaching and success in gaining and sustaining employment among homeless people. *Research on Social Work Practice*, 26: 668–674.

Howlett, M. A., McWilliams, M. A., Rademacher, K., O'Neill, J. C., Maitland, T. L., Abels, K., Demetriou, C., & Panter, A. T. 2021. Investigating the effects of academic coaching on college students' metacognition. *Innovative Higher Education*, 46: 189–204.

Ianiro, P. M., & Kauffeld, S. 2014. Take care what you bring with you: How coaches' mood and interpersonal behaviour affect coaching success. *Consulting Psychology Journal*, 66: 231–257.

Ianiro, P. M., Lehmann-Willenbrock, N., & Kauffeld, S. 2015. Coaches and clients in action: A sequential analysis of interpersonal coach and client behaviour. *Journal of Business and Psychology*, 30: 435–456.

Ianiro, P. M., Schermuly, C. C., & Kauffeld, S. 2013. Why interpersonal dominance and affiliation matter: An interaction analysis of the coach-client relationship. *Coaching*, 6: 25–46.

IntHout, J., Ioannidis, J., & Borm, G. F. 2014. The Hartung–Knapp–Sidik–Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian–Laird method. *BMC Medical Research Methodology*, 14. doi: 10.1186/1471-2288-14-25

Jones, R. J., Woods, S. A., & Guillaume, Y. 2015. The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, 89: 249–277.

Junker, S., Pömmer, M., & Traut-Mattausch, E. 2021. The impact of cognitive-behavioural stress management coaching on changes in cognitive appraisal and the stress response: A field experiment. *Coaching*, 14: 184–201.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. 2012. Publication bias in the organizational sciences. *Organizational Research Methods*, 15: 624–662.

Kim, S., & Kuo, M. 2015. Examining the relationships among coaching, trustworthiness, and role behaviors: A social exchange perspective. *Journal of Applied Behavioral Science*, 51: 152–176.

Kochanowski, S., Seifert, C. F., & Yukl, G. 2010. Using executive coaching to enhance the effects of behavioural feedback to managers. *Journal of Leadership & Organizational Studies*, 17: 363–369.

Koole, S. L., & Tschacher, W. 2016. Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance. *Frontiers in Psychology*, 7. doi: 10.3389/fpsyg.2016.00862

LeBlanc, P. J. 2018. Higher education in a VUCA world. *Change: The Magazine of Higher Learning*, 50: 23–26.

Liljenstrand, A. M., & Nebeker, D. M. 2008. Coaching services: A look at coaches, clients, and practices. *Consulting Psychology Journal*, 60: 57–77.

Losch, S., Traut-Mattausch, E., Mühlberger, M. D., & Jonas, E. 2016. Comparing the effectiveness of individual coaching, self-coaching, and group training: How leadership makes the difference. *Frontiers in Psychology*, 7. doi: 10.3389/fpsyg.2016.00629

McGonagle, A. K., Beatty, J. E., & Joffe, R. 2014. Coaching for workers with chronic illness: Evaluating an intervention. *Journal of Occupational Health Psychology*, 19: 385–398.

McGonagle, A. K., Schwab, L., Yahanda, N., Duskey, H., Gertz, N., Prior, L., Roy, M., & Kriegel, G. 2020. Coaching for primary care physician well-being: A randomized trial and follow-up analysis. *Journal of Occupational Health Psychology*, 25: 297–314.

Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. 2004. A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, 72: 1050–1062.

Morris, S. B. 2008. Estimating effect sizes from pretest–posttest–control group designs. *Organizational Research Methods*, 11: 364–386.

Niglio de Figueiredo, M., Krippeit, L., Ihorst, G., Sattel, H., Bylund, C. L., Joos, A., Bengel, J., Lahmann, C., Fritzsche, K., & Wuensch, A. 2018. ComOn-Coaching: The effect of a varied number of coaching sessions on transfer into clinical practice following communication skills training in oncology—results of a randomized controlled trial. *PLoS One*, 13. doi: 10.1371/journal.pone.0205315

Passmore, J., & Rahman, H. 2012. Coaching as a learning methodology: A mixed methods study in driver development—a randomized controlled trial and thematic analysis. *International Coaching Psychology Review*, 7: 166–184.

Passmore, J., & Velez, M. J. 2012. Coaching fleet drivers: A randomized controlled trial (RCT) of short coaching' interventions to improve driver safety in fleet drivers. *Coaching Psychologist*, 8: 20–26.

Poepsel, M. 2011. ***The impact of an online evidence-based coaching program on goal striving, subjective well-being, and level of hope*** (Doctoral dissertation, Capella University). Available from ProQuest dissertations and theses database. (UMI No. 3456769)

Poluka, L. A., & Kaifi, B. A. 2015. Performance coaching within the telecommunications industry. ***Journal of Applied Management and Entrepreneurship***, 20: 49–65.

Ramseyer, F., & Tschacher, W. 2011. Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. ***Journal of Consulting and Clinical Psychology***, 79: 284–295.

Sbarra, D. A., & Hazan, C. 2008. Coregulation, dysregulation, self-regulation: An integrative analysis and empirical agenda for understanding adult attachment, separation, loss, and recovery. ***Personality and Social Psychology Review***, 12: 141–167.

Schwarzer, G. 2007. Meta: An R package for meta-analysis. ***R News***, 7: 40–45.

Sidik, K., & Jonkman, J. N. 2002. A simple confidence interval for meta-analysis. ***Statistics in Medicine***, 21: 3153–3159.

Singh, P., Aggarwal, R., Tahir, M., Pucher, P. H., & Darzi, A. 2015. A randomized controlled study to evaluate the role of video-based coaching in training laparoscopic skills. ***Annals of Surgery***, 261: 862–869.

Sonesh, S. C., Coultas, C. W., Lacerenza, C. N., Marlow, S. L., Benishek, L. E., & Salas, E. 2015. The power of coaching: A meta-analytic investigation. ***Coaching***, 8: 73–95.

Spence, G. B., & Grant, A. M. 2007. Professional and peer life coaching and the enhancement of goal striving and well-being: An exploratory study. ***Journal of Positive Psychology***, 2: 185–194.

Stelter, R., Nielsen, G., & Wikman, J. M. 2011. Narrative-collaborative group coaching develops social capital–a randomised control trial and further implications of the social impact of the intervention. ***Coaching***, 4: 123–137.

Stiles, W. B., Barkham, M., & Wheeler, S. 2015. Duration of psychological therapy: Relation to recovery and improvement rates in UK routine practice. ***British Journal of Psychiatry***, 207: 115–122.

Stiles, W. B., Honos-Webb, L., & Surko, M. 1998. Responsiveness in psychotherapy. ***Clinical Psychology: Science and Practice***, 5: 439–458.

Stoeger, H., Balestrini, D. P., & Ziegler, A. 2021. Key issues in professionalizing mentoring practices. ***Annals of the New York Academy of Sciences***, 1483: 5–18.

Sue-Chan, C., & Latham, G. P. 2004. The relative effectiveness of external, peer and self-coaches. ***Applied Psychology***, 53: 260–278.

Taie, E. S. 2011. Coaching as an approach to enhance performance. ***Journal for Quality and Participation***, 34: 34–38.

Taylor, L. M. 1997. ***The relation between resilience, coaching, coping skills training, and perceived stress during a career threatening milestone*** (Doctoral dissertation, Georgia State University). DAI-B 58/05, p. 2738, Nov. 1997.

Tee, D., Shearer, D., & Roderique-Davies, G. 2017. The client as active ingredient: "Core self-evaluations" as predictors of coaching outcome variance. ***International Coaching Psychology Review***, 12: 125–132.

Telle, N. T., Moock, J., Heuchert, S., Schulte, V., Rossler, W., & Kawohl, W. 2016. Job maintenance through supported employment PLUS: A randomized controlled trial. ***Frontiers in Public Health***, 4: 194.

Theeboom, T., Beersma, B., & van Vianen, A. E. M. 2014. Does coaching work? A meta-analysis on the effects of coaching on individual level outcomes in an organizational context. ***Journal of Positive Psychology***, 9: 1–18.

Toegel, G., & Nicholson, N. 2005. Multisource feedback, coaching, and leadership development: personality and homophily effects. ***Academy of Management Proceedings***, 2005. doi: 10.5465/ambpp.2005.18779257

Ungerer, C., Heinzelmann, N., Baltes, G. H., & König, M. 2019. ***The effect of business coaching on NTBF survival: Findings and lessons learned from a randomized controlled trial.*** In *2019 IEEE international conference on engineering, technology and innovation (ICE/ITMC)*: 1–10. doi: 10.1109/ICE.2019.8792604

Vecchione, M., Alessandri, G., & Marsicano, G. 2014. Academic motivation predicts educational attainment: Does gender make a difference? ***Learning and Individual Differences***, 32: 124–131.

Viechtbauer, W. 2010. Conducting meta-analyses in R with the Metafor package. ***Journal of Statistical Software***, 36: 1–48.

Viering, S., Jäger, M., Bärtsch, B., Nordt, C., Rössler, W., Warnke, I., & Kawohl, W. 2015. Supported employment for the reintegration of disability pensioners with mental illnesses: A randomized controlled trial. ***Frontiers in Public Health***, 3. doi: 10.3389/fpubh.2015.00237

Wasylyshyn, K. M. 2022. The "art" of executive coaching at the top: Using clients' self-imagery as a tool for high impact. ***Consulting Psychology Journal***, 74: 1–18.

Watson, J. C., & Wiseman, H. E. 2021. ***The responsive psychotherapist: Attuning to clients in the moment***. Washington, DC: American Psychological Association.

Wieder, G., & Wiltshire, T. J. 2020. Investigating coregulation of emotional arousal during exposure-based CBT using vocal encoding and actor–partner interdependence models. ***Journal of Counseling Psychology***, 67: 337–348.

Williams, J. S., & Lowman, R. L. 2018. The efficacy of executive coaching: An empirical investigation of two approaches using random assignment and a switching-replications design. ***Consulting Psychology Journal***, 70: 227–249.

Yip, J., Trainor, L. L., Black, H., Soto-Torres, L., & Reichard, R. J. 2020. Coaching new leaders: A relational process of integrating multiple identities. ***Academy of Management Learning & Education***, 19: 503–520.

Zanchetta, M., Junker, S., Wolf, A. M., & Traut-Mattausch, E. 2020. "Overcoming the fear that haunts your success": The effectiveness of interventions for reducing the impostor phenomenon. ***Frontiers in Psychology***, 11. doi: 10.3389/fpsyg.2020.00405

Zimmermann, L. C., & Antoni, C. H. 2020. Activating clients' resources influences coaching satisfaction via occupational self-efficacy and satisfaction of needs. ***Zeitschrift für Arbeits- und Organisationspsychologie***, 64: 149–169.

————————⋀⋀————————

Erik de Haan (erik.dehaan@ashridge.org.uk) is the director of the Ashridge Centre for Coaching, Hult International Business School, and professor of organization development and coaching at VU University Amsterdam. He has an MSc in theoretical physics, an MA in psychodynamic psychotherapy, and a PhD in psychophysics, and specializes in team coaching and one-to-one coaching for executives.

Viktor O. Nilsson (viktor.nilsson@ashridge.hult.edu) is an adjunct at Hult International Business School and a research data manager for administrative data in the Centre for Experimental Research on Fairness, Inequality and Rationality at the Norwegian School of Economics. He completed his MSc in quantitative research methods at University of Strathclyde, Glasgow.

————————⋀⋀————————