The barriers to a scientific, evidence-based coaching evaluation practice

Miriam Schneider[1], Almuth McDowall[2] and David Tee[3]

[1]CoachHub, Berlin
[2]Birkbeck, University of London
[3]CoachHub, London

Abstract

Workplace coaching has grown in popularity and is increasingly used for a plethora of purposes across organisations. As a growing field, coaching is still in need of a continuing commitment to evidence-based evaluation, especially considering the current unsystematic outcome literature. However, this need for scientific, evidence-based evaluation is not actioned and there is indication that coaching evaluation is even less rigorous in practice. This position paper explores what might be the barriers against a scientific, evidence-based coaching evaluation in practice. Suggestions grounded in the literature are presented with the aim that these might inform future research and practice.

Keywords: coaching evaluation; evidence-based practice, evidence-based evaluation

Introduction

Workplace coaching has become a favoured leadership development approach and a multibillion-dollar industry, with over 70,000 practitioners worldwide (ICF, 2020). Coaching has developed as a practitioner-driven field where the research evidence is still catching up (Passmore & Lai, 2019). This is not helped by an ongoing coaching definition debate which points to inherent challenges for building a coherent knowledge base (Passmore & Tee, 2020). These challenges include a plethora of applications and outcomes of coaching (Grant, 2013; Jones et al. 2016), the heterogeneity of coaching methodologies (Grant, 2013) and a lack of process research about how coaching works (Athanasopoulou & Dopson, 2018). Coaching as a complex, dynamic process (Ely et al., 2010) may be hard to capture solely through standardised scientific evaluation methods (Osatuke et al., 2017).

Nonetheless, meta-analyses have generated evidence that coaching 'works' though with varying effects (Theeboom et al., 2013; Jones et al., 2016) on different outcomes. Yet concern remains about a disjointed coaching outcome literature (Jones & Underhill, 2018). As early as 2007, Grant and Cavanagh lamented a lack of well-designed evaluation studies, the need for a common understanding of outcomes, more randomised controlled studies and more longitudinal research. Over a decade later, there remains unclear agreement on universally accepted outcomes and key determinants for coaching effectiveness (de Haan & Duckworth, 2013; Bozer & Jones, 2018; Jones & Underhill, 2018). Furthermore, Grant (2013) outlines that a majority of evaluation research has been practitioner-led and cautions about the validity of many of the findings, as best practices have not always been incorporated into all evaluation studies, such as many small-scale qualitative studies lacking methodological and theoretical rigour. There is also criticism of existent coaching evaluation research is an exaggerated focus on quantitative outcome research, which gives too little

recognition to the complexity of coaching and its 'active ingredients'. 'Active ingredients' are variables argued to influence coaching effectiveness and as moderating and mediating factors are argued to be critical to coaching success (de Haan et al., 2013; Bachkirova et al., 2015).

Grant and Cavanagh (2007) argue that the key to evidence-based practice (EBP) in general, which should also apply to a thriving coaching discipline, is an evaluation approach grounded in science and research. Considering this advocacy for more EBP, the question arises what stops practitioners from conducting scientific, evidence-based coaching evaluation (EBE). There is a gulf between academic calls for rigour and scientific evaluation studies and what is done in practice. The purpose of this position paper, therefore, is to review the literature which explores EBE with a focus on actual and potential barriers. We commence by framing the meaning of an 'evidence-based' approach to coaching and the case for evidence-based practice. We will then critically review various EBE approaches as outlined by the literature and what they indicate regarding barriers to a more rigorous evaluation approach.

Evidence-based practice

The notion of EBP, in which scientific insights aim to inform professional practice, finds its origins in the medical field (Briner & Rousseau, 2011). Driven especially by a quantitative, positivist paradigm, randomised controlled trials (RCTs) have been positioned as the gold standard for informing the most effective medical treatments. Briner and Rousseau (ibid.) have translated this concept to organisational psychology, outlining four sources of information that should be considered for an EBP: "Practitioner expertise and judgment, evidence from the local context, a critical evaluation of the best available research evidence, and the perspectives of those people who might be affected" (p.6). They state the goal of EBP is not to find the 'perfect answer' but to consciously make informed and improved decisions (ibid.). The scientist-practitioner model by Hodgkinson (2006) has also informed such views upon EBP and it speaks to the importance of theoretical and methodological accuracy, being specific about insights from findings as well as what could not be discovered. He further mentions the importance of having knowledge and skills on how to evaluate the impact of interventions and to the importance of sufficient understanding about "[…] core concepts, theories, tools, and techniques that constitute the field, and understand the principal research methods […]" (p.176).

It is posited that there are seven identified necessary characteristics of EBP in the organisational psychology field: the knowledge of the term evidence-based, the existence and accessibility of research findings and summaries in general as well as to practitioners, a healthy scepticism about fashionable fads, a demand for EBP from decision makers, integrative decision making based on the four sources outlined above and continued professional training on EBP (Briner & Rousseau, 2011). The authors further critique the limited extent to which these characteristics are considered in organisational psychology.

However, beyond these views as to what EBP constitutes, there is also a debate about criteria as to what and who determines the best available evidence, with Cassell (2011) cautioning that evidence by itself is "politics, values and interest laden" (p.23). Cassell further argues that systematic reviews, posited by Briner and Rousseau (2011) as 'good' evidence, generally subscribe to positivist research methodologies and that there are other sources of legitimate evidence beyond just systematic reviews or randomised controlled trials. In summary, a debate as to what constitutes EBP, its lacking application in practice as well as some critical points of view keep the space of EBP on its toes.

We argue that, despite this debate, the understanding of the evidence-base for coaching brings value to practice and credibility as coach practitioners (Grant & O'Connor, 2019). Furthermore, "an evidence-based approach is the foundation on which our future success will be built, and the yardstick against which it will ultimately be measured" (Linley, in Grant & Cavanagh, 2007, p.252). Some experts advocate for double-blind randomised controlled trials (RCTs) to avoid coaching languishing as "alternative medicine" (Furnham, in Grant & Cavanagh, 2007, p.242). A need for "collecting rigorous empirical evidence" (Sheldon, in Grant & Cavanagh, 2007, p.242) is advocated and thereby being better able to "demonstrate return on investment" (Clutterbuck, in Grant & Cavanagh, 2007, p.242). This call for evidence continues to date, albeit a little "broader and less reductionist" (Grant, 2016a, p.76). Indeed, in terms of EBE, Grant (2016) argues specifically against the medical model's hierarchy of evidence, with its preference for RCTs and almost total disregard towards qualitative research, as this limited approach does not match the reality of what informs and shapes the richness within coaching. This also aligns with broader criticisms of EBP as discussed earlier. He instead argues that both quantitative and qualitative research, if "well-designed and peer-reviewed" (p.80) constitute valuable evidence given that coaching is inherently different from medicine. In the latter we are interested if a drug works and what may be any potential side effects. In coaching, the question is when and for whom the intervention works - and with what kind of activities and individuals. As there is this strong call for EBP from academics in coaching psychology, we do now need to turn to the barriers which may inhibit its wide scale adoption.

Barriers to evidence-based practice

The first barrier to EBP is the apparent lack of demand for rigorously determined, quantitative impact data from organisations (Briner & Rousseau, 2011). This is particularly interesting as it contrasts with the strong call for EBP from science. This distinction may be driven in part by the differing goals practitioners and academic researchers wish to achieve with research (Grant, 2016). The lack of demand may also be due to EBE requiring sufficient resource and knowledge. Thus, it needs the willingness of managers and decision makers to engage in such a process and the knowledge of stakeholders as to how to execute the evaluation data generation and analysis. However, management tends to aim for short-term results and return on investment (ROI) (Briner & Rousseau, 2011), so EBP's focus on improving "the process and outcome of decision making" (Briner & Rousseau, 2011, p.19) may not give the answers organisational management wishes to hear, nor within the timescale they prefer.

Second, there are some concerns about EBP perpetuating a preference for quantitative research. Although the importance of qualitative evidence is outwardly promoted, some researchers voice concern over its authenticity. Additionally, as mentioned, there is debate about criteria as to how evidence is determined to be valuable in the first place (Cassell, 2011). For a practice such as coaching, which is conducted in applied settings and has many unobservable aspects (such as meaning making and identity) central to its effectiveness, quantitative research methods have a legitimate role, but not necessarily an exclusive one. Therefore, coaching research does contain a wealth of non-quantitative research, with alternative designs used to legitimately explore topics such as the critical moments in coaching (de Haan, 2019). The question which arises around these different forms of evidence is one about the preference of types of evidence in the practical setting of an organisation.

The third barrier to EBP is the lack of recognition by researchers of politics and power as a form of evidence beyond the scientific (Bartlett, 2011). Especially in relation to evaluation, those with power may decide what is desirable evidence and whether resources for EBE are made available. Bartlett (2011) argues that power and politics are an important part of evidence, especially in organisational psychology, proposing that the four-source model outlined by Briner and Rousseau (2011) should include questions about politics and power. Indeed, Ben-Hador's (2016) research on coaching being perceived as a "tacit performance evaluation" (p.75) emphasises the importance of politics being a form of evidence in coaching. Most EBP approaches may disregard the macro level of wider societal discourses, described as the "institutionalised patterns of knowledge and power" (Western, 2017, p.42). Western (2017) raises an interesting point by highlighting the "dominance of managerial scientific-rationalism in the workplace" (p. 48), which he argues influences the above discussed scientific EBP. This lack of criticality in EBP is echoed by Shoukry and Cox's (2018) opinion piece on coaching as a social process and Bachkirova and Borrington's (2020) cautioning of coaching becoming a 'beautiful idea' that makes us ill, as it may be (ab)used to put the onus of betterment upon the individual and take responsibility away from the organisation. These critical explorations are rarely discussed in organisational settings.

Considering the literature on EBP, it seems as though a lack of demand and skills, a focus on positivism, aspects of power and politics and a lack of criticality may, individually and in combination, act as barriers to an evidence-based practice and indeed evaluation.

Evidence-based coaching evaluation

We now turn to the scientific, evidence-based evaluation of coaching. The debate about evaluating organisational development interventions has been a constant ever since Kirkpatrick (1959) introduced his evaluation model. Although it is outside the focus of this paper, it is contextually relevant to state that there also continues to be a wider debate about training evaluation in general (Passmore & Velez, 2012).

Coaching evaluation has been recognised as a difficult feat (Bachkirova et al., 2015; Jones & Underhill, 2018), and it is argued that the theory and methodology is somewhat unsystematic and underdeveloped (Peterson & Kraiger, 2004; Greif, 2013; Jones & Underhill, 2018). Furthermore, academics and practitioners have different approaches to research and evaluation. Among these tensions is the desire by practitioners for real world benefit and impact, whereas researchers aim to contribute to the knowledge base and theory (Grant, 2016). De Haan and Duckworth (2013) capture this tension when stating that, ethically, the researcher-practitioner needs to prioritise fulfilling their coaching commitment, even if in doing so, they might disrupt the clinically detached studying of effectiveness by deviating from standardised aspects of an experimental design. Nevertheless, there are several approaches which can inform EBE in practice.

Grant (2013) asserts that the key ambition of evaluation in practice is to evaluate the effects of coaching to make better investment decisions. Peterson and Kraiger (2004) propose two further motives: marketing to gather internal support or promote coaching externally, and programme improvement. These three purposes can be subsumed under summative and formative evaluation approaches. The former evaluates outcomes and effects of a coaching intervention. It can be both qualitative and quantitative. However, arguably, the quantitative approach has been the focus in coaching research (Bachkirova & et al., 2015; Ely et al., 2010). Formative evaluation supports the development of programme designs, based on the best available evidence, evaluating a programme throughout to ensure success (Ely et al.,

2010; Greif, 2013). As Ely and colleagues (2010) outline, both approaches are important to coaching as it is a complex, dynamic and context-related intervention. It is surprising that no one to date referenced ethical and moral grounds for evaluation: human-to-human coaching (as distinct from, say, AI bot coaching) is contingent on close relationships and a high level of trust. As de Haan (2021b) highlights, coaching could also do potential harm if either or both are absent. Nevertheless, this paper's final section will focus on summative evaluation, as it is here that other barriers to EBE may be located.

Summative evaluation

The desire in outcome studies to determine whether coaching works (Fillery-Travis & Lane, 2006) has been the focus of much of coaching research in the opening decades of this century (Passmore & Lai, 2019). Thus, it is not surprising that coaching evaluation has mostly subscribed to a positivist paradigm (Grant, 2013; Bachkirova et al., 2015). Outcome studies in this paradigm vary from qualitative case studies (e.g. Freedman & Perry, 2010) to quantitative, large-scale RCTs (de Haan, 2021a). Indeed, there are now several meta-analyses - statistical summaries of the research field - that speak to the effectiveness of coaching. For instance, DeMeuse et al. (2009), Theeboom et al. (2013), Jones et al. (2016) and Wang et al. (2021) all found positive effects of coaching, though with varying effect sizes. Furthermore, they encountered the challenge that the studies within their scope explored a variety of different dependent variables. This lack of a shared research agenda and consensus on variables of interest continues to challenge EBE (Jones & Underhill, 2018).

However, variation in outcomes is part of the complex nature of coaching (Greif, 2013; Jones & Underhill, 2018). Jones et al. (2016) developed a criteria taxonomy based on Kirkpatrick's (1959) and Kraiger et al.'s (1993) training evaluation models to tackle the breadth of dependent variables in coaching impact studies, separating findings into affective, cognitive or skill-based learning outcomes and individual, team and organisational results. Whether this taxonomy is widely adopted to bring focus and more easily aggregated data is yet to be seen, not least as arguably other outcomes, such as wellbeing, are just as important particularly in the present context (Wang, Lai, Bao & McDowall, 2021).

Grant (2013) proposed three rigorous study designs for EBE to conduct outcome evaluation: namely case studies, within-participant research, and between-participant RCTs. He subsequently cautions that, although (qualitative) case studies provide a complex picture of coaching's effectiveness, they cannot permit generalisation or comparisons of results. Therefore, there continues to be a need for quantitative evaluation, as argued by many proponents of EBE (Peterson & Kraiger, 2004; Greif, 2013; Jones & Underhill, 2018). Whilst both within-participant research and between-participant studies can provide more generalisable and quantitative insights, most coaching evaluation studies to date have used a within-participant research design, comparing a group's pre- and post-assessment outcomes (de Haan, 2021a). However, within-subject research does not allow for the attribution of changes directly to coaching or any ability to make causal inferences. Nevertheless, Grant (2013) argues that well designed within-subject research does permit the use of inferential statistics to make predictions regarding coaching impact.

Between-participant and RCT studies do allow us to make causal attributions (de Haan, 2021a). However, they also require a large group of participants, separated into one intervention and one control condition. To achieve RCT standards in practice, such as a genuinely random allocation to conditions, the researcher requires a large enough number of participants in each group (around 60), as well as ideally a placebo intervention. As Passmore

and Fillery-Travis (2011) and de Haan (2021a) argue, these design requirements hit barriers in real world settings, such as businesses with real, time-sensitive staff development requirements. An additional barrier concerns the withholding of coaching from staff in the control condition, which may prove politically and ethically problematic in practice (Peterson & Kraiger, 2004). Thus, while RCTs are regarded as the gold-standard in the medical field, they may not be such in coaching (Grant, 2013) as coaching is arguably contextual and not devoid of context or framing.

The literature provides several detailed recommendations on how to conduct EBE in practice. Peterson and Kraiger (2004), for example, suggest a comprehensive five step guide that directs the evaluation process from deciding on its purpose to designing the evaluation process, its implementation, analysis and the communication of findings. However, they give little guidance on appropriate measures. More recently, Jones and Underhill (2018) offer a detailed overview of such measures and recommend appropriate tools for self-rated and multi-source feedback based on the criteria taxonomy developed by Jones et al. (2016). They further outline the challenges related to deploying many measures, in that they typically give little to no insight regarding team and organisational level results. Furthermore, Ely and colleagues (2010) emphasise the importance of multi-source, multi-level, and longitudinal evaluation as some coaching outcomes may only be realised many months after the coaching intervention has concluded (Ely et al., 2010; Greif, 2013).

Barriers to summative evaluation

Despite much knowledge in the summative coaching evaluation field, there are also challenges which echo the barriers identified to EBP in general. The first is the state of the delineation of the coaching field as outlined at the start. Additionally, coaches not only come from varied backgrounds such as psychology, business, or adult education (DeMeuse et al., 2009), but also use varying approaches, making comparability across cases difficult (Jones et al., 2016).

Second, the focus of coaching can vary enormously, covering topics from stress and emotional intelligence to leadership development, talent retention and organisational responses to macro-economic pressures. Unsurprisingly therefore, the measures used to evaluate coaching are similarly varied (Grant, 2013) and, to date, there is no single agreed outcome measure for coaching (Grover & Furnham, 2016; Jones & Underhill, 2018) which reflects the plethora of coaching foci and approaches. This further extends to the ROI debate. Whilst it is difficult to determine organisation-level outcomes of coaching (Jones et al., 2016), some researchers (e.g. McGovern et al., 2001; Phillips, 2007) have used ROI calculations in their studies. Although this is a commendable effort to close the outcome expectation gap for organisational leaders, such approaches are open to much criticism (Grant, 2012) due to their reliance on estimates from coaching clients which may be biased by these individuals' own interests within the organisation (Athanasopolou & Dopson, 2018; Grant, 2012; Peterson & Kraiger, 2004).

The third barrier to EBE, regardless of the source of any impact data, concerns an over-reliance on self-reported data (Theeboom et al, 2013) which, especially in retrospective studies, faces distortions such as hindsight bias or reduction of cognitive dissonance (DeMeuse et al., 2009). The client may not remember their starting point accurately after a long coaching programme and would also have probably invested time and effort into the coaching, thus wanting to confirm its value. EBE researchers may need to seek a middle ground between capturing the perceptions of the client and triangulating outcomes by, for

example, reporting feedback from other sources such as team members or managers (Athanasopoulou & Dopson, 2018). The use of validated scales can further reduce these biases (Jones & Underhill, 2018), with Greif (2013) recommending 'goal attainment scaling' as an evaluation option that may even serve as a unified outcome measure. As outlined by Briner and Rousseau (2011) earlier, all the above methodological challenges may be heightened by a lack of evaluation and analytic skills of those responsible for the process.

Coaches may also be asked to provide evaluation. Although this can serve both the client and the organisation, the coach may face a conflict of interest. Indeed, Greif (2015) suggests some coaches may oppose EBE in practice, both for fear of bad reviews and due to the actual or perceived impracticality of the scientific approach. He provocatively stated that "there is nothing more to fear for charlatans than a scientific evaluation" (p.176, free translation).

Finally, knowledge about the methodological approaches and theories of evaluation is not often widespread in practice. Statistical knowledge, understanding of and access to validated scales and how to use these as well as rigorous qualitative methodologies may not be as easily accessible to practitioners as would be expected in research. In summary, while there are many calls for an EBE, there are also many barriers that may hinder this in practice.

Implications of these barriers for research and practice

An interesting piece of research on the use of EBP, has found that "client demands and acceptability to the client are more frequently considered than evidence from the scientific research literature and […] that such client concerns "trump" scientific research evidence" (Bartlett & Francis-Smythe, 2016, pp. 621-622). This indicates that, in addition to all the barriers to an evidence-based practice in general as well as the inherent barriers to an evidence-based coaching evaluation specifically, power may play a much bigger role than anticipated in practising an evidence-based coaching evaluation and partly addresses Hodgkinson's question as to why EBP may not be used as rigorously as hoped by researchers.

The need of EBP to be appropriately resourced and a seemingly disproportionately strong demand for quantitative data may point to conflict of resources within organisations and serve as a further barrier to EBE. Interestingly, these two barriers play into the concern around power and politics and the question as to who in an organisation determines 'what is good evidence' and how much permission is given to be critical of any decisions about that.

In summary, EBP faces a series of barriers already, with more added when evaluation specifically is considered. Going back to the specifications of EBP, we would expect practitioners who embark on an evidence-based coaching evaluation to have the knowledge and skills, to adhere to methodological and theoretical rigour, to understand the principles of research methods and to have access to and understanding of scientific and practitioner knowledge. Even if this were given for every practitioner embarking on an EBE journey, they are still facing barriers inherent to coaching evaluation itself. The variety of coaching methodologies, and a multiplicity of outcomes for coaching pose inherent challenges to evaluation; how can we measure the impact of an outcome which we do not know we need to measure? And if we were to measure every imaginable outcome, how is this practicable in a work environment? An over-reliance on asking the coaching client to evaluate their coaching success and other methodological challenges add to the list of barriers. And finally, there may be a conflict of interest. This may come from the coaches who are asked to evaluate the

coaching. As they will need to sell their service, they may find themselves in a conflict when they need to raise an issue of the coaching process and evaluate the impact thereof. In addition, this conflict of interest around EBE may also come up for coach commissioners or leaders in the organisation who are closely connected in the process of purchase, the desire for success and the input they have committed.

Implications for research from the findings of this review are to further explore context of formative coaching evaluation. Additionally, it might be interesting to empirically explore the experiences of coaching practitioners or buyers when faced with the decision of evidence-based coaching evaluation. This would enable a more in-depth understanding of what might overcome barriers to it. Or to say it in the words of Hodgkinson (2011):"to gain a more detailed and systematic understanding of why the demand for nonevidence-based practices and solutions so often outstrips the demand for evidence-based ones" (p.52). Further research could explore the impact of power and politics on coaching evaluation. Who decides on what constitutes good evidence and how might that decision be influenced? Does it vary across sizes and cultures of organisations? Another area of research might want to shine a light on the skills and self-efficacy of practitioners around the EBE process in practice. What are their biggest pain points and how might they be alleviated? And finally, beyond barriers, it would be of interest to identify enablers to an evidence-based coaching evaluation as well so that the field of coaching can benefit from more widely used evidence-based practice in coaching evaluation.

Conclusion

This review contributed to the debate of the science-practitioner gap by reviewing the barriers to evidence-based practice in organisational psychology with a special focus on evidence-based coaching evaluation. Several barriers from both evidence-based practice as well as coaching evaluation have been identified – this may be a case for the age-old saying 'better the devil you know' as, by starting to understand the barriers, one may start to explore how to overcome them. And it is high time as over a decade ago there was already an "unequivocal consensus for the need for an evidence-based approach to coaching" (Grant & Cavanagh, 2007, p.1) which can ensure the continued flourishing of the coaching field.

References

Athanasopoulou, A., & Dopson, S. (2018). A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most? *The Leadership Quarterly*, *29*(1), 70-88.

Bachkirova, T., Arthur, L., & Reading, E. (2015). Evaluating a coaching and mentoring programme: Challenges and solutions. *International Coaching Psychology Review*, *10*(2), 175-189.

Bachkirova, T., & Borrington, S. (2020). Beautiful ideas that can make us ill: Implications for coaching. *Philosophy of Coaching: An International Journal*, *5*(1), 9-30.

Bartlett, D. (2011). The neglect of the political: An alternative evidence-based practice for I–O psychology. *Industrial and Organizational Psychology*, *4*(1), 27-31.

Bartlett, D., & Francis-Smythe, J. (2016). Bridging the divide in work and organizational psychology: evidence from practice. European Journal of Work and Organizational Psychology, 25(5), 615-630.

Ben-Hador, B. (2016). Coaching executives as tacit performance evaluation: A multiple case study. *Journal of Management Development*, *35*(1), 75-88.

Bozer, G., & Jones, R. J. (2018). Understanding the factors that determine workplace coaching effectiveness: A systematic literature review. *European Journal of Work and Organizational Psychology*, *27*(3), 342-361.

Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I–O psychology: Not there yet. *Industrial and Organizational Psychology*, *4*(1), 3-22.

Cassell, C. (2011). Evidence-based I–O psychology: what do we lose on the way? *Industrial and Organizational Psychology*, *4*(1), 23-26.

de Haan, E. (2019). *Critical moments in executive coaching: Understanding the coaching process through research and evidence-based theory* (1st ed.). Routledge/Taylor & Francis Group.

de Haan, E. (2021a). *What Works in Executive Coaching: Understanding Outcomes Through Quantitative Research and Practice-Based Evidence* (1st ed.). Routledge.

de Haan, E. (2021b). The case against coaching. *The Coaching Psychologist, 17*(1), 7-13.

de Haan, E., Duckworth, A., Birch, D., & Jones, C. (2013). Executive coaching outcome research: The contribution of common factors such as relationship, personality match, and self-efficacy. *Consulting Psychology Journal: Practice and Research*, *65*(1), 40.

de Meuse, K. P., Dai, G., & Lee, R. J. (2009). Evaluating the effectiveness of executive coaching: Beyond ROI? *Coaching: An international journal of theory, research and practice*, *2*(2), 117-134.

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. *The leadership quarterly*, *21*(4), 585-599.

Fillery-Travis, A., & Lane, D. (2006). Does coaching work or are we asking the wrong question? *International Coaching Psychology Review, 1(1), 23-36.*

Freedman, A.M. and Perry, J.A. (2010). Executive consulting under pressure: A case study. *Consulting Psychology Journal: Practice and Research*, *62*(3), 189-202.

Grant, A. M. (2012). ROI is a poor measure of coaching success: towards a more holistic approach using a well-being and engagement framework. *Coaching: An International Journal of Theory, Research and Practice*, *5*(2), 74-85.

Grant, A. M. (2013). The efficacy of coaching. In J. Passmore, D. B. Peterson, & T. Freire (Eds.), *The Wiley-Blackwell handbook of the psychology of coaching and mentoring* (pp. 15-39). Wiley Blackwell.

Grant, A. M. (2016). What constitutes evidence-based coaching? A two-by-two framework for distinguishing strong from weak evidence for coaching. *International journal of evidence-based coaching and mentoring*, *14*(1), 74-85.

Grant, A. M., & Cavanagh, M. J. (2007). Evidence-based coaching: Flourishing or languishing? *Australian Psychologist*, *42*(4), 239-254.

Grant, A., & O'Connor, S. (2019). A brief primer for those new to coaching research and evidence-based practice. *The Coaching Psychologist*, *15*(1), 3.

Greif, S. (2013). Conducting organizational-based evaluations of coaching and mentoring programs. In J. Passmore, D. B. Peterson, & T. Freire (Eds.), *The Wiley-Blackwell handbook of the psychology of coaching and mentoring* (pp. 445-470). Wiley Blackwell.

Greif, S. (2015). Wie wirksam ist Coaching? Ein umfassendes Evaluationsmodell für Praxis und Forschung. In R. H. Wegener, A. Fritze, & M. Loebbert (Eds.), *Coaching-Praxisfelder. Praxis und Forschung im Dialog* (2. Aufl. ed., pp. 159-178). Wiesbaden: Springer VS.

Grover, S., & Furnham, A. (2016). Coaching as a developmental intervention in organisations: A systematic review of its effectiveness and the mechanisms underlying it. *PloS one*, *11*(7), e0159137.

Hodgkinson, G. (2006). Commentary: The role of JOOP (and other scientific journals) in bridging the practitioner– researcher divide in industrial, work and organizational psychology (IWO) psychology. Journal of Occupational and Organizational Psychology, *79*, 173-178. doi:10.1348/096317906X104013

Hodgkinson, G. P. (2011). Why evidence-based practice in I–O psychology is not there yet: Going beyond systematic reviews. *Industrial and Organizational Psychology*, *4*(1), 49-53.

International Coaching Federation & PricewaterhouseCoopers LLP. (2020). 2020 ICF global coaching study executive summary. Accessed at URL: https://coachfederation.org/app/uploads/2020/09/FINAL_ICF_GCS2020_ExecutiveSummary.pdf.

Jones, R. J., & Underhill, B. O. (2018). Mastering evaluation. In J. Passmore, B. Underhill & M. Goldsmith (Eds.), *Mastering Executive Coaching* (pp. 41-58). Abingdon, Oxon: Routledge.

Jones, R. J., Woods, S. A., & Guillaume, Y. R. F. (2016). The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. *Journal of Occupational and Organizational Psychology*, *89*(2), 249-277.

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs: Reaction'. *Training and Development Journal, 18*, 3-9.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of applied psychology*, *78*(2), 311-328.

McGovern, J., Lindemann, M., Vergara, M., Murphy, S., Barker, L., & Warrenfeltz, R. (2001). Maximizing the impact of executive coaching. *The Manchester Review*, *6*(1), 1-9.

Osatuke, K., Yanovsky, B., & Ramsel, D. (2017). Executive coaching: New framework for evaluation. *Consulting Psychology Journal: Practice and Research*, *69*(3), 172-186.

Passmore, J., & Fillery-Travis, A. (2011). A critical review of executive coaching research: a decade of progress and what's to come. Coaching: An International Journal of Theory, Research and Practice, 4(2), 70-88.

Passmore, J., & Lai, Y.-L. (2019). Coaching psychology: Exploring definitions and research contribution to practice? *International Coaching Psychology Review*, 14(2), 69–83.

Passmore, J. & Tee, D. (2020). Insights from the Nature of Coaching Psychology and Coaching Supervision. In Passmore, J. & Tee, D. (Eds.), *Coaching researched: A coaching psychology reader* (pp.1-2). John Wiley & Sons Ltd.

Passmore, J., & Velez, M. (2012). SOAP-M: a training evaluation model for HR. *Industrial and Commercial Training*, *44*, 315-325.

Peterson, D. B., & Kraiger, K. (2004). A practical guide to evaluating coaching: Translating state-of-the-art techniques to the real world. In Edwards, J.E., Scott, J.C., & Raju, N.S. (Eds.), *The human resources program evaluation handbook,* (pp.262-282). Thousand Oaks, CA:Sage.

Phillips, J.J. (2007). Measuring the ROI of a coaching intervention, Part 2. *Performance Improvement*, *46*, 10-23.

Shoukry, H., & Cox, E. (2018). Coaching as a social process. *Management Learning*, *49*(4), 413-428.

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

Wang, Q., Lai, Y. L., Xu, X., & McDowall, A. (2021). The effectiveness of workplace coaching: a meta-analysis of contemporary psychologically informed coaching approaches. Journal of Work-Applied Management.

Western, S. (2017). The key discourses of coaching. In T. Bachkirova, G. Spence, & D. Drake (Eds.), *The SAGE handbook of coaching* (pp. 42-61). Sage Publications, Inc.