The Leadership Quarterly 29 (2018) 70–88

Contents lists available at ScienceDirect

The Leadership Quarterly

journal homepage: www.elsevier.com/locate/leaqua

A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most?

Andromachi Athanasopoulou[a,b,*], Sue Dopson[b]

[a] School of Business and Management, Queen Mary University of London, UK
[b] Saïd Business School, University of Oxford, UK

ARTICLE INFO

Keywords:
Executive coaching
Systematic review
Social context
Research designs

ABSTRACT

In this article, we focus on a specific type of personal and professional development practice -executive coaching- and present the most extensive systematic review of executive coaching outcome studies published in peer-reviewed scholarly journals to date. We focus only on coaching provided by external coaches to organizational members. Our purpose is twofold: First, to present and evaluate how executive coaching outcome studies are designed and researched (particularly regarding methodological rigor and context-sensitivity). Secondly, to provide a comprehensive review of what we know about executive coaching outcomes, what are the contextual drivers that affect coaching interventions and what the current gaps in our understanding of coaching practice. On that basis, we discuss and provide a research agenda that might significantly shift the field. We argue that methodological rigor is as important as context-sensitivity in the design of executive coaching outcome studies. We conclude with a discussion of implications for practice.

Introduction

Within less than three decades of existence, the executive coaching (EC) field has rapidly grown to become a multibillion-dollar global market (Armstrong, 2011, p. 183). As of 2012 there were approximately 47,500 professional coaches worldwide with nearly $2 billion total revenue generated by coaching globally (ICF, 2012). Over time, the nature of coaching interventions has significantly changed. A Harvard Business Review survey of 140 coaches revealed that just over a decade ago coaches were mostly hired to address toxic behaviors in leadership, whereas now they are hired to develop high-potential performers (Coutu et al., 2009, p. 92), including to assist coachees' transition to new roles (Sherpa Coaching Survey, 2014). These developments have affected management education. Courses that foster reflection and personal development are becoming popular in MBA curricula and executive education portfolios (Petriglieri, Wood, & Petriglieri, 2011; Datar, Garvin & Cullen, 2010).

On the other hand, despite its high demand, the coaching industry still seeks professional legitimacy and is seen as a developing field with high variation in coaches' background, coaching practices and quality (Drake, 2008; Ennis et al., 2008; ICF, 2014). An International Coach Federation study found that the profession's biggest obstacles are "untrained coaches" and confusion in the marketplace about coaching benefits (ICF, 2012). This lack of clarity is also reflected in research. The field still lacks a "clear and agreed sense" of what "outcomes" should be or how they should be measured (De Haan & Duckworth, 2013, p. 12).

Despite its short history, the EC field has produced a small number of review papers that have sought to survey the research and practice of coaching (e.g. Jones, Woods, & Guillaume, 2016; Grover & Furnham, 2016; Theeboom, Beersma & van Vianen, 2014; Segers, Vloeberghs, Henderickx & Inceoglu, 2011; Ely et al., 2010; Feldman & Lankau, 2005; Kampa-Kokesch & Anderson, 2001). These papers and our work are in line with recent calls (Arbaugh, 2011; Rynes & Brown, 2011) for more review-type pieces in management education and learning. Building on prior research (Athanasopoulou & Dopson, 2015), we offer a systematic review of all peer-reviewed articles on EC outcomes and discuss the research and practice implications. We argue that although methodological rigor is important and discussed in prior meta-analyses and other review studies on EC outcomes, the social contextual aspects of a coaching intervention have been largely neglected in such reviews. We, therefore, call for a reframing of the future research agenda that takes these into account.

This is the first study that systematically reviews in such depth both the "what" (coaching impact and quality of evidence) and the "how" and "why" (coaching practice and social contextual influences) of EC. The field has been preoccupied with *whether* coaching works and has paid much less attention to *how* it works. It has been mostly focused on

* Corresponding author at: Queen Mary University of London, School of Business & Management, 4th Floor, Francis Bancroft Building, Mile End Road, London E1 4NS, UK.
E-mail addresses: a.athanasopoulou@qmul.ac.uk, andromache.athanasopoulou@sbs.ox.ac.uk (A. Athanasopoulou), sue.dopson@sbs.ox.ac.uk (S. Dopson).

https://doi.org/10.1016/j.leaqua.2017.11.004
Received 4 February 2017; Received in revised form 24 November 2017; Accepted 26 November 2017
Available online 03 January 2018
1048-9843/

a micro-level analysis of EC outcomes, which is unsurprising considering the one-on-one nature of coaching practice. EC research has also been indirectly preoccupied with the *meso* level (how changing one's behavior or improving one's leadership skills leads to better interactions with individuals and groups within and outside the organization) and very limitedly with the macro level (organizational benefits from EC). Research questions that link or integrate these (micro, *meso* and macro) levels of contextual analysis warrant the field's attention, too.

What is executive coaching?

In this study, we focus on EC provided by an external to the organization coach typically in collaboration with the organization, excluding all other coaching or consulting practices (e.g. life coaching, internal/managerial coaching, mentoring). EC is a *targeted, purposeful intervention* that helps executives develop and maintain positive change in their personal development and leadership behavior (Grant, 2012a). As such, it is a "process" which involves the partnership of three key stakeholders: the coach, the coachee (i.e. the executive) and the coachee's sponsoring organization (Ennis et al., 2008; Garman, Whiston & Zlatoper, 2000; Kampa-Kokesch & Anderson, 2001; Kilburg, 1996; Michelman, 2004; O'Neill, 2007; Witherspoon & White, 1996). It is different from counseling and other therapeutic interventions (Passmore, 2009, p. 272). Unlike psychotherapy, EC does not address mental health problems (De Haan, Duckworth, Birch & Jones, 2013) and unlike counseling and psychotherapy where performance measurement is primarily based on client self-report, EC measurement should relate to the executive's and sponsoring organization's bottom-line performance (Kampa-Kokesch & Anderson, 2001, p. 211). Most importantly -and different to other interventions- the individual goals of the intervention must "always link back and be subordinated" to strategic organizational objectives (Ennis et al., 2008, p. 23).

The coaching intervention is characterized by high context-sensitivity as a result of the unique mix of environments, characteristics, motivations and attitudes of stakeholders who have direct effects on coaching outcomes. Moreover, the diversity of coaches' backgrounds and training (e.g. business, psychology, sports) brings variations to coaching practices employed (Bono, Purvanova, Towler, & Peterson, 2009). Under such a pluralistic conceptual and practice backdrop, we set out to conduct a systematic review of the field.

Research questions and data collection process

EC outcome research is young. Kampa-Kokesch and Anderson (2001, p. 206) found only seven empirical studies up to 2000 that explored the effectiveness of EC [Foster & Lendl (1996), Garman et al. (2000), Gegner (1997), Hall, Otazo & Hollenbeck (1999), Judge & Cowell (1997), Laske (1999) and Olivero, Bane & Kopelman (1997); later amended in an article erratum by Kampa-Kokesch and Anderson to include Peterson's (1993a) dissertation]. As of 2005, Feldman and Lankau (2005, p. 830) identified "fewer than 20 studies that have investigated executive coaching with systematic qualitative and/or quantitative methods". In a more recent review, Ely et al. (2010) identified 49 leadership coaching evaluation studies (only 20 peer-reviewed, with the rest being non-peer-reviewed, dissertation and conference presentations). We identified 110 peer-reviewed outcome studies on executive coaching, with 32 of them published in journals with an impact factor.

Review studies – including ours- agree that despite the significant growth of EC outcome studies over the last 20 years, the research quality varies. A challenge is that coaching studies are often carried out by practitioners who may pay little attention -if at all- to carefully crafted research procedures. On the other hand, more scholars now use experimental or quasi-experimental methods which are particularly promising for outcome evaluation (e.g. Osatuke, Yanovsky, & Ramsel, 2016; Bozer, Sarros & Santora, 2014; Moen & Federici, 2012a, 2012b; Grant, Green & Rynsaardt, 2010). This echoes the field's call for evidence-based coaching that draws on behavioral and social sciences (Grant, 2003; Stober & Grant, 2006) to increase its credibility and quality of practice (Drake, 2009, p. 12). It also reflects the need to take stock of the wide range of outcome studies produced so far.

Why this review of the field - and why now?

Our focus on EC outcome research complements Segers et al.'s (2011) conceptual effort to understand the coaching industry via a 3-dimensional theoretical framework [coaching agendas (i.e. what); coaches' characteristics (i.e. who); and coaching approaches/schools (i.e. how)]. We believe that improvements in research quality will help the industry to enhance its status as a profession built on evidence-based practices. Our work both complements and is different from prior reviews of EC outcomes. Among the most notable efforts to review the field are the early qualitative reviews by Kampa-Kokesch and Anderson (2001) and Feldman and Lankau (2005) and the more recent systematic reviews and meta-analyses by Ely et al. (2010), Theeboom et al. (2014), Jones et al. (2016) and Grover and Furnham (2016). We next explain how our work is distinctive compared to these and where we make a contribution.

Kampa-Kokesch and Anderson's (2001) and Feldman and Lankau's (2005) early reviews may not be as in-depth regarding the various aspects of EC outcome research as our study and the four other systematic reviews and meta-analyses are. Yet, these were the field's early attempts to take stock of its research. Kampa-Kokesch and Anderson (2001) offered the first, seminal review of this literature. Later, Feldman and Lankau (2005) sought to summarize key outcome studies, research methods and content issues regarding the coaching practice (e.g. coaching relationship and coaching approaches) and proposed a new research agenda. Our review is distinctively different from the field's four main reviews in recent years (Ely et al., 2010; Theeboom, Beersma & van Vianen, 2014; Jones, Woods & Guillaume, 2016; Grover & Furnham, 2016): Ely et al. (2010) systematically reviewed -as we did- both qualitative and quantitative studies, but they focus specifically on leadership coaching evaluation. The authors distinguish between formative and summative evaluations and offer a detailed review of methodologies, data sources, analysis approaches, and evaluation criteria. In comparison, we not only systematically review these elements in all (110) peer-reviewed EC outcome studies published until the end of 2016, *but also assess a wider range of variables beyond the evaluation parameters* explored by Ely et al. (2010). Moreover, we review studies for their contextual considerations, positive or negative outcomes, how these relate to each coaching stakeholder and whether social context has been accounted for.

Besides broadening our analysis on methodological issues relating to these studies, the focus on *social context* is a key contribution that we make here and one that none of the prior review studies and meta-analyses have effectively explored. For instance, Jones et al.'s (2016, p. 254) statistical meta-analysis sought to take a multi-level approach (individual, team, and organizational outcomes) and explore coaching evaluation criteria at the individual level across affective, cognitive and skill-based outcomes. Their intention was to test whether the reviewed outcome studies converge in that workplace coaching works and at which levels or outcome categories the effects are strongest. By comparison, we do not seek to quantify the effectiveness of EC, but instead shed light on weaknesses of the research designs used and discuss the need for a more context-sensitive research approach. Theeboom et al. (2014) also conducted a statistical meta-analysis of only 18 quantitative studies but focused only at the individual level, providing a numerical value for the positive effect coaching has.

A further distinction in our work is that -unlike Jones et al.'s (2016) and Theeboom et al.'s (2014) meta-analyses and Grover and Furnham's (2016) (qualitative) systematic review- we examine the full set of EC

outcome studies, not a subset. These three studies, for instance, focus only on quantitative studies [i.e. within-subjects and between-subjects research designs for Jones et al. (2016), empirical and practitioner quantitative control group and non-control studies on executive, leadership and business coaching forGrover and Furnham (2016) and heavily screened quantitative studies forTheeboom et al. (2014)]. We also do not fully agree with some of the level classifications in Grover and Furnham's (2016, p. 18) review when, for instance, they assign transformational leadership, performance -as rated by others- and manager behavior as organizational level outcomes. We would classify them as individual level outcomes. What we also find in both Jones et al.'s (2016) and Grover and Furnham's (2016) well-executed work is an attempt to look into mechanisms that underlie coaching effectiveness but only based on quantitative studies. Yet, since such mechanisms are located at the social context of the intervention, they cannot be fully captured by the quantitative studies alone. Indeed, Grover and Furnham (2016, p. 26) admit that qualitative studies are particularly crucial for EC research and while they did not include them in their review they do note (p. 35) that a key gap future research needs to address is a focus on contextual aspects such as cultural, age or gender differences and coachee's personality. Similarly, Theeboom et al. (2014, p. 14) called for a shift in the research from "'does it work?' to 'how does it work?'" to help identify the underlying mechanisms and processes of coaching. Last but not least, our study is the first systematic review that also explores whether there are *any reported outcomes relating to coaches*. We seek to make a timely contribution to the field via the most thorough and context-sensitive review of the "what", "how" and "why" of EC across all existing peer-reviewed studies, irrespective of the research methods each employed.

Our research questions

It could be argued that it is too soon to focus on the context of EC considering the major methodological issues that prior meta-analyses and our work has identified. Yet, we feel that the quality of research goes hand-in-hand with the type of research questions we ask. We contend that there has been too much emphasis in the literature on the EC outcomes at the expense of the processes or contextual factors that affect these outcomes. A review of the EC outcome literature is incomplete unless these are addressed, too. No other systematic review has sought to explore how EC research treats EC practice as independent of or embedded in a broader social context. We have two research questions that serve four objectives:

Research Question 1: *How* are EC outcomes researched and *what* are the strengths and weaknesses of their research designs?

Objective 1a: Present and evaluate how these studies are designed and researched.

Objective 1b: Discuss the strengths and weaknesses from the use of different research methods or designs in future EC outcome studies and associated implications.

Research Question 2: *What* do we know about EC outcomes and *how* and *why* do contextual factors affect these outcomes?

Objective 2a: Examine what evidence exists within the EC research regarding the outcomes of EC and the factors that affect these outcomes as well as identify gaps in reported evidence.

Objective 2b: Explore the implications on research and practice based on this evidence.

To fully address each research objective, we conducted first- and second-order analyses per research question. Objectives 1a and 2a are addressed in the Data Findings section and Objectives 1b and 2b are addressed in the Implications section.

Data collection and analysis: the process and the data

Our work is a systematic review described as interpretation meta-synthesis (Hoon, 2013) focusing as much on "what" as on "how" and "why" (see Pawson et al.'s (2005) 'realist' approach to systematic review). Unlike (quantitative) meta-analysis (i.e. aggregation synthesis) which is "rooted in the positivist tradition" with knowledge "conventionally aggregated" in "time- and context-free generalizations", *interpretation meta-synthesis* is appropriate for studying an individual or a process (Hoon, 2013, pp. 524 and 526), hence, a perfect analysis vehicle for studying the EC practice and its stakeholders. The theoretical contribution emerges by taking the studies' local contexts into account (Hoon, 2013). Hoon (2013) employs this method for case study research alone, however we use it to review and interpret outcomes from studies using various qualitative and/or quantitative research methods. We adjusted Hoon's (2013, p. 529) protocol of eight meta-synthesis steps as: 1. Framing of research questions, 2. Locating relevant research, 3. Extracting and coding data, 4. Analyzing data (i.e. identifying the reported outcomes and the variables affecting these outcomes), 6. Synthesizing outcomes and variables (i.e. accumulating both outcomes and variables into broader categories and mapping which variables relate to which stakeholders), 7. Building theory (by focusing on the role of context and human agency in shaping the outcomes and identifying gaps in the literature) and 8. Discussing the meta-synthesis results.

Screening process

We started by reviewing Grant's (2011) annotated bibliography of the abstracts of all scholarly publications on executive, workplace and life coaching to that date and continued with further systematic review on PsycINFO and Business Source Complete to identify any additional EC outcome studies published up until December 2016. After screening for only the studies that focus on some kind of EC outcome research, we cross-checked our list across Ulrichsweb database to identify articles appearing only in peer-reviewed journals. This led to a list of 110 articles published in 37 peer-reviewed journals. After further screening based on the Journal Citation Reports (JCR 2013 and 2016 Social Science and Science Editions) from the Web of Science database we found that just over half of these journals (21 of 37) are listed in Journal Citation Reports as having an impact factor with less than a third of peer-reviewed EC articles (32) published in these journals. The 16 journals that have published the remaining 78 of the 110 peer-reviewed EC papers are not listed in JCR as journals with an impact factor. Since with our first research question we seek to evaluate how all EC outcome studies are researched, we decided to focus on all 110 peer-reviewed papers, rather than the subset of papers that have an impact factor. Therefore, in response to this research question we focused on all 110 peer-reviewed papers. However, in order to address our second research question, we had to screen for studies based on the assessment of question 1. This meant that from the initial set of 110 peer-reviewed studies discussed in research question 1, we then selected only those studies that were designed and executed in a way that offered confidence for their validity. Therefore, we excluded studies where the author was the coach (i.e. lacked research independence) and the two studies that calculated 'Return on Investment' (ROI) for coaching (Parker-Wilkins, 2006; Phillips, 2007). The lack of research independence is one of the main shortcomings of existing EC outcome research which will be explained later in our review as we discuss the methodological challenges associated with EC research. We also excluded the two ROI studies identified in our initial list because none offers confidence regarding rigorous calculations of ROI. These studies displayed large difference in EC ROI calculations with one estimating it to 689% (Parker-Wilkins, 2006) and the other to 221% (Phillips, 2007). This signals inconsistency in calculations, but we also felt that both had

weaknesses in their data collection method. For instance, Parker-Wilkins (2006) relied on interviewees' estimate of the percentage of monetary value for each benefit attributed to EC (i.e. value creation of one or more of eight business impact areas). Phillips' (2007) method of data collection was more rigorous and drew on multiple sources of data, combining hard data (sales growth, direct cost reduction or retention of key staff) to intangible benefits (increased commitment or reduced stress), converting them into monetary value based mainly on in-house experts' or coachees' estimates. Again, these calculations drew mainly on estimates.

The exclusion of the research dependent studies and the two ROI studies led to a shortlist of 84 studies. We discuss this process later in the paper. We also conducted a more focused review on the subset of articles with an impact factor to check whether the quality of this subset is significantly higher than the rest of the non-impact factor but peer-reviewed studies. We, also, discuss separately key issues relating to the screened-out studies.

Collected data

Our initial 110-study review led to a 68-page, single-spaced groundwork table where we mapped each study across six dimensions: i. authors' names, ii. the study theme (i.e. what each study seeks to examine/what the key research question is) iii. Data collection/research methodology, iv. executive coaching approach/method used in the study, v. key findings and vi. publication source. The data on the *theme/research question* allowed us to explore the themes covered in existing outcome studies (i.e. each study's research question) and whether their focus is on the coach, coachee, sponsoring organization or the coaching process. From the *research methodology* dimension we identified the design and methods used in outcome studies, which and how many studies have used each research method and whether the author was the coach (research dependence). From the *coaching approach/method* dimension we identified how many and which studies referred to a specific coaching approach or method. In the *findings* dimension we examined whether the reported outcomes are "positive", "negative", "conflicting/moderate" (e.g. a partially successful intervention, not producing the intended positive effect) according to measurement criteria set out by the authors of each study. We also noted when the outcome study was just "descriptive", when the focus was on describing aspects of the EC practice rather than changes resulting from an intervention (e.g. De Haan and Nieß' (2012) study of "critical moments" in coaching is a descriptive study).

Finally, we examined whether the *social context* of the intervention was addressed in the analysis or discussion of each study. This work is summarised in Table 1, for the preparation of which, we revisited studies and double-checked the accuracy of the extended table entries.

For question 2 (i.e. what we know about EC outcomes) we wanted to get qualitative information regarding: a. the types of outcomes EC produces and evidence of potential pitfalls (see Table 2) and b. what existing EC outcome studies report as factors that influence the outcome of an intervention (see Table 3).

We developed Table 2 by gathering first separately all positive outcomes and all "conflicting/moderate" or "negative" outcomes reported in the shortlisted 84 studies (i.e. excluding research dependent and ROI studies). We identified more than 70 positive outcomes, which we grouped into 11 broad categories (see the first column inTable 2 for categories of positive outcomes and the second column for a sample of such outcomes). In Table 2, we organized these 11 categories across the three coaching stakeholders these related to. For instance, overcoming regressive behaviors relates to the coachee, whereas positive perceptions of coach's effectiveness relate to the coach. We also sought evidence of coaching pitfalls as drawn by the evidence on moderate/conflicting or negative EC outcomes. We found only 16 studies that report some kind of not fully-positive outcome that can be grouped into eight types of pitfalls (some observed in more than one study). These pitfalls are presented in Table 2 (last column) as potential challenges practitioners must keep in mind.

We also wanted to see what factors affect the EC outcomes. We identified these factors by reviewing our *study theme* and *key study findings* dataset and then revisiting each study. We searched for factors that the authors had set out to test in terms of their potential impact on outcomes. For example, Lewis-Duarte and Bligh (2012) examined coaches' perceived use and effectiveness of proactive influence tactics. They found that influence tactics were more frequently associated with client commitment and contributed to positive outcomes. Thus, we registered "coaches' use of influence tactics" as a factor influencing EC outcomes. We present the full list of factors in Table 3, thematically grouped across the intervention, the organization, the coachee, the coach and the relationship among stakeholders. In the last column of Table 3 we summarize aspects or issues relating to these factors that remain unexplored or understudied. This became the basis for our Implications for Research section.

Assessing the quality of executive coaching outcome research (Q1a)

In this section, we respond to research question 1 (Objective 1a; Evaluate how these studies are designed and researched). The coaching outcomes research is characterized by high heterogeneity of issues, problems and goals that are selected as the focus of the intervention (Greif, 2007). This makes the comparison of outcomes across studies difficult. Some studies explore the effectiveness of a specific coaching method (e.g. Foster & Lendl, 1996), other studies present case studies of individual coachees (Peterson & Millier, 2005) and others look into case studies of groups of coachees within organizations (Moen & Federici, 2012a; Lawrence, 2015; Ben-Hador, 2016). Then, there are studies that examine outcomes regarding specific intervention elements (e.g. De Haan et al.'s (2013) and Smith and Brummel's (2013) studies on EC active ingredients) and other studies focus on coaches' effectiveness (e.g. Nikolova, Clegg, Fox, Bjørkeng & Pitsis, 2013).

We found that most studies (69) have the coachee either as their sole focus (37 studies) or in combination with other stakeholders (32 studies), while the coach is the sole focus in fewer studies (18) (Table 1). We were surprised to see that only three studies focus on organizational-level outcomes. Also, the fact that only four studies (e.g. Winum, 2005; Levenson, 2009) focused on all three stakeholders (organization, coachee, coach) confirms that EC outcome research erroneously treats coaching as an individual-level intervention rather than a social process with active involvement of multiple stakeholders. Finally, no study so far has looked into the coach-sponsoring organization engagements, that is, the quality and nature of the contracting stage of EC between the coach and the sponsoring organization and the effect this has on coaching outcomes.

Coaching approaches used

Less than half of the studies (39) mention the coaching method used (e.g. psychodynamic, solution-focused, GROW; Table 1). The cognitive-behavioral approach, followed by the solution-focused and positive psychology/strengths coaching approaches are the most frequently reported approaches in EC outcome studies. Recent studies are more likely to report the type of coaching method used (e.g. MacKie, 2014; Zarecky, 2014; Howard, 2015). While a quantitative meta-analysis of the effectiveness of each coaching approach is beyond the scope of this study, knowing how each approach works and what its produced outcomes are, could help the field refine these approaches or better understand under what circumstances they work best. Unfortunately, we did not find any study that compares outcomes from different coaching methods. Instead, a few studies compared the effect of EC to other types of developmental interventions (Olivero, Bane & Kopelman, 1997; Sue-Chan & Latham, 2004; Moen & Skaalvik, 2009; Jones, Woods & Guillaume, 2016) or to no intervention (e.g. Grant et al., 2010; Moen &

Table 1
Key research characteristics of peer-reviewed executive coaching outcome studies (up to December 2016) (N = 110).

1a: Research methodology N[a]																	
Case study - individual executive(s) or coaches	18	Case study – organization(s)	26	Survey/ quantitative (only)	17	Qualitative interviews or transcript (only)	16	Mixed methods/ types of data	37	Experimental/quasi-exp./randomized controlled studies	15	Meta-analysis or systematic review of emp. studies	5	Return on investment (ROI)	2		
1b: Executive coaching approach or framework/model reported N[a]																	
Cognitive-behavioral/ cognitive/ behavioral	9	Positive psychology/ strengths coaching	5	Solution-focused	5	Emotional Intelligence; Systems-oriented approach; goal-setting; GROW	3 each	Gestalt; NLP; psycho-dynamic, developmental coaching	2 each	Competency approach; coaching in context; salutogenisis; transformational coaching; eye movement desensitization and reprocessing (EMDR); three principles psychology approach; process consultation coaching	1 each	Mixed approaches and/or models	6	Not specified or not applicable	71		
1c: Focus of study (stakeholder or process focused?) N																	
Coachee	37	Coach	18	Sponsoring organization	3	Coach & coachee interaction	12	Coachee & sponsoring organization	16	Coach & sponsoring organization	0	All 3 stakeholders	4	Focus on the intervention as a process	20		
1d: Context considered in data analysis and/or discussion? N																	
Yes (sociological consideration of context)					50		Descriptive			33		No (acontextual studies)			27		
1e: Are the outcomes positive? N[a]																	
Positive					75		Moderate/conflicting or negative			16		Descriptive			32		
1f: Research independence: (i.e. author is not the coach in the study) N																	
Yes					86					No		24					

[a] Numbers do not always add up to 110 because some studies may belong to more than one category (e.g. use of multiple coaching approaches or reporting a mix of positive and negative outcomes).

Table 2
Evidence of positive executive coaching outcomes and potential pitfalls according to the outcome studies [N = 84; 110 peer-reviewed studies minus 24 author-as-coach (research dependent) studies and 2 ROI studies].

Categories of positive outcomes from coaching	Examples of specific positive outcomes per category	Evidence of potential pitfalls for practice based on outcome studies findings
A. The coachee		
a. Personal development		
1. Overcoming regressive behaviors or experiences	- Reduced stress/anxiety	*1. Is coaching effective for everyone?* In cases of severe executive derailment coaching can be ineffective. In some groups coaching more effective than others.
2. EC seen as effective, positive or life-changing experience	- Coaching seen as source of support and encouragement - Increased work and life satisfaction	*2. Who defines what counts as an effective coaching intervention?* For coaching purchasers (e.g. HR department) coaching success means positive business results and for coachees it means personal development benefits
3. Better personal management/self-control	- Improved resilience - Better time-management	
4. Improved personal skills/abilities or acquisition of new ones	- Adaptability/flexibility - Improved ability and quality of goal-setting	*3. Openness to coaching experience?* Coachees do not always see the value of EC
b. The coachee & the others: behavioral changes in relation to others		
5. Better leadership skills	- Better management and development of others - Coachee perceived by others (seniors and subordinates) as a more effective leader post-coaching	
6. Better quality of interactions and relationships	- Improved team player & team-building skills - Better communication skills	
c. The coachee & his/her work		
7. Work performance/productivity and planning	- Positive impact on psychological variables that affect work performance (e.g. self-awareness, development of authentic leadership qualities) - Improved coachees' understanding, fit, relatedness and commitment to sponsoring organization - Improved agenda setting skills	*4. Can coaching-induced goal-setting have negative effects on work-performance?* Coaching may contribute to a (self-reported) negative relationship between goal-achievement and productivity/work performance when coachees set too difficult goals or devote less time to accomplish them or have low pre-coaching motivation
8. Nurturing working environment	- Feeling more valued at work - Better ability to build cross-functional relationships - Enhanced workplace well-being	
B. The organization		
9. Positive organizational-level outcomes	- Indirect positive organizational effects from increased employee satisfaction, productivity, leadership effectiveness and coaching culture	*5. Is it worth to invest on coaching?* Variations among outcome studies on how significant the difference is in behavior between a coached and a non-coached group *6. Fitting coaching to context?* In small-medium-sized organizations the intervention impact is stronger on personal attributes than business attributes
C. The coach		
10. Positive perceptions of coach's effectiveness	- Eliciting coachee's commitment to the intervention	*7. Coaches rarely measure own performance so as to improve their practice* Mostly rely on informal feedback
11. Coach's personal development	- Coaches experience of self-actualisation and fulfilment - Obtained new knowledge and improved coach skillset	8. Coping with resistance to coaching e.g. Difficulty to convince some managers on the importance of some aspects of the practice.

Federici, 2012a, 2012b; MacKie, 2014).

Research methodologies – strengths and weaknesses

We found EC peer-reviewed studies to have large variations in methodological robustness and triangulation of findings (Denzin, 1984; Stake, 1995) - an observation also confirmed by Ely et al. (2010). Case studies are the most popular method and there is a growing use of mixed methods. Nearly half (44) of the studies are case studies of individual coachees or of coaching interventions for several executives in an organization. Case study research can be a source of rich data both about outcomes and about contextual drivers. However, we found that many EC case studies simply consist of narratives of the coaching intervention for one or more coachees (e.g. Kiel, Rimmer, Williams and Doyle's (1996) study of a sales and marketing executive or Winum's (2005) study of a high-potential African American executive). While single coachee case studies have value as outcome vignettes, they are not generalizable unless a more context-sensitive analysis is pursued such as De Haan and Nieß' (2012) study of critical moments in coaching. Rigorous research typically involves multiple case studies and cross-case analysis (Eisenhardt, 1989). This is an area where EC outcome research lags: 11 studies present coaching vignettes of one or more coachees (e.g. Foster & Lendl, 1996; Anderson, 2002), but do not look for patterns or variations across cases.

Another method questioned about its potential to measure outcomes (Grant, 2012b) is the Return on Investment (ROI), with only two peer-reviewed EC studies having used this method (Parker-Wilkins, 2006; Phillips, 2007). For reasons described earlier, both of those studies were excluded from our analysis. A problem with ROI is that it assigns monetary value onto every outcome, based on estimates, which can be arbitrary and filtered by organizational politics and personal egos. One has self-interest to attribute high monetary value onto their own contribution to organizational outcomes and this would inflate the ROI calculations. Even if triangulated with data from multiple sources, ROI is the conflation of several contextual factors (e.g. personal contribution, power and influence on organizational processes, market conditions, business cycle). The coaching contribution to any improvements cannot be easily isolated (Wise & Voss, 2002). Grover and Furnham (2016) observed that no study has experimentally calculated the coaching ROI, implying that ROI may not be as useful a measure of EC outcomes at the organizational level – to that effect, we agree. According to Phillips (2007, p. 15) there are more "credible isolation methods", such as control groups and trend analysis, but these have not been used in ROI calculations.

A promising development, particularly over the last five years, is the use of mixed methods such as combinations of surveys with qualitative interviews, comparison of pre-and post-coaching assessments, which are found in a third (37) of the reviewed 110 studies (Table 1). Only a

Table 3
Factors affecting coaching outcomes: what do existing outcome studies indicate and how can research help?
(**Key**: Randomized Controlled Studies = RCSs; Interviews = I; Surveys = S; Case Studies = CS; Multisource Feedback and other Personality Assessment Tools = MF).
Factors marked with an * have been tested in outcome studies with experimental designs.

Factors affecting outcomes	What works according to existing outcome studies	Gaps & recommended research method/data collection
THE INTERVENTION		
Coaching model*	- Every coaching model tested brought positive outcomes	- Compare coaching models' level of effectiveness (*Method*: *RCSs*)
Occurrence of critical moments	- Issue-related or self-related "new realizations" (e.g. self-doubt, new learning) evoke positive emotions	- How can coaches increase or induce critical moments? (*RCSs*; *CS*; *I*; *S*)
Use of personality assessment tools*	- Coach's timely and effective use of assessment tools improves coaching effectiveness and vice versa	- Compare personality assessment tools to test their effectiveness in coaching interventions (*RCSs*; *S*; *MF*)
Intervention focus and implementation	- Focus on coachee's strengths, on learning and/or developing goal orientation (*with caution*) and grounding to action plan	- How can a coach facilitate the alignment of coaching stakeholders' goals? (*I*; *CS*; *RCSs*)
Use of influence tactics	- Coach's use of influence tactics (e.g. coalition, rational persuasion) increases coachee commitment	- Are some influence tactics more effective on certain coachees or to elicit certain outcomes? (*CS*; *RCSs*; *I*; *S*)
Language and communication	- Coach mimicking client's language helps intervention outcomes	- Is such mimicking helpful in any organizational context? (*CS*; *RCSs*)
Coaching setting, duration and means*	- Long-term coaching more effective than short-term - Telephone coaching as powerful coaching tool	- What makes outcomes sustainable? - Compare coaching outcomes within exec ed. and organizational settings (*RCSs*; *CS*; *MF*)
THE ORGANIZATION		
Organizational support	- Coaching signals employer's support to coachee; whether support is real or perceived, it improves coaching impact	- How can an organization best select which executives to sponsor for coaching to maximize its organizational benefits? (*I*; *S*; *MF*)
Integration of coaching to leadership development	- It signals the organization's commitment to coachee's development	-How can a leadership development programme more effectively integrate coaching to make its outcomes sustainable? (*RCSs*; *CS*; *MF*)
Size and type of organization/ industry	- In SMEs coaching impact is stronger on personal attributes than on business attributes	- Are some coaching models better for certain industries or organizations? (*Method*: *RCSs*; *CS*)
Organizational culture	- A supportive organizational environment/culture contributes to coaching success	- What attributes does a supportive organizational culture have? (*I*; *S*)
THE COACHEE		
Coachee's personal attributes	- Openness to experience, conscientiousness, self-awareness, extraversion, maturation and emotional stability, self-efficacy and confidence	- How to identify executives with these attributes or find ways to develop them - Are some individuals "uncoachable"? (*I*; *S*; *MF*)
Coachee's expectations of outcomes*	-Coachees' self-efficacy beliefs and positive outcome expectancies increase likelihood of success	- Are intrinsic or extrinsic parameters contributing to these expectancies? (*RCSs*; *CS*; *I*; *S*; *MF*)
Sensemaking of intervention	- Perceived supervisor support and sense of personal developability	- In-depth study of coachee's sensemaking of coaching experience (*CS*; *I*)
Coachee's learning style*	- Self-directed learning and learning goal orientation help, provided there is pre-coaching motivation	-Which learning styles are more conducive to positive coaching outcomes? (*RCSs*; *CS*; *I*; *S*; *MF*)
Pre-, during and post-coaching motivation*	- Motivation to be coached and to transfer skills to work and commitment to coaching	- Explore techniques coaches can use to increase or sustain coachee's motivation (*CS*; *RCSs*; *I*; *S*)
Job rank*	- Impact on performance is stronger for middle managers and their subordinates than for executives	-How should senior leaders' coaching be different from middle managers'? (*CS*; *RCSs*; *I*; *S*; *MF*)
THE COACH		
Coach's background	- Coach selection and coaches' intervention style differs depending on coaches' background	- Compare styles and outcomes between coaches with psychology and business backgrounds (*RCSs*; *S*; *I*; *MF*)
Coach's behavior, skills, abilities and quality of practice*	- Positive state of mind, authenticity, active listening and empathy, reflective questioning, learning and development facilitation, agility and ability to understand context and interpret results and ethical challenges and professionalism	- What attributes make a good coach and how to develop them? (*I*; *S*)
THE RELATIONSHIP AMONG COACHING STAKEHOLDERS		
Coach-coachee relationship and fit	- Good coach-coachee fit and relationship-building (including mutual trust, honesty and respect)	- What coach and coachee attributes should coaching purchasers look for when deciding on coach-coachee dyads? (*RCSs*; *CS*; *MF*)
Stakeholder alignment and collaboration	- Contracting, alignment on purpose, methods, measurement criteria and desired outcomes	- What procedures and coach intervention designs facilitate stakeholder alignment? (*I*; *S*; *CS*)
Coachee's and coach's gender	- Impact of coach-coachee dyads' gender similarity on EC effectiveness and the impact of gender on coach selection	- Are male or female coaches better? Are same gender coach-coachee dyads more successful? (*CS*; *RCSs*; *MF*)
Clarity in roles & expectations	- Agreeing on stakeholders' roles and responsibilities and contracting outcomes facilitates success	- What should a good coaching engagement contract include? (*I*; *S*)

few studies use solely qualitative data –mainly interviews (16)- or surveys (17).

The quasi-experimental or experimental/randomized controlled studies (RCSs) have recently gained traction within the coaching outcome studies and are growing in popularity (e.g. Grant et al., 2009; Moen & Federici, 2012a and 2012b; Osatuke, Yanovsky, & Ramsel, 2016) (15 of 110). Well-crafted RCSs can be particularly helpful in testing the effects of EC.

Finally, five studies have used meta-analysis or systematic review of all types of EC or of a specific intervention (e.g. Ward, Van de Loo & ten Have's (2014) meta-analysis focused on the psychodynamic approach only).

Self-reporting of outcomes and research independence

A key consideration when assessing the robustness of reported outcomes is who reports them. Although collecting data from multiple sources is plausible, it is not always feasible and this remains a challenge for any statistical model, including leadership mediation models

(Fischer, Dietz & Antonakis, 2017, p. 1741). Consistent with earlier findings by Ely et al.'s (2010) and Grover and Furnham's (2016) reviews, we observed that EC outcome studies significantly rely only on coachees' and/or coaches' self-reporting (e.g. Tamir & Finfer, 2016; Gray, Burls, & Kogan, 2014; Van Diemen van Thor (2014); Zarecky (2014); Bowles & Picano, 2006; Dawdy, 2004; Wales, 2003). We identified nearly a quarter of studies (24 of 110) where the author is the coach of the intervention. Coaches' self-reporting of outcomes compromises the research independence. A key concern is that data based on self-reports result in inflation of the self-assessment of one's performance (Kruger & Dunning, 1999; Peterson, 1993b) and hence, of EC outcomes (Theeboom et al., 2014 and Ely et al., 2010). This applies both when the coach is the author of a study and when the coachee self-reports outcomes. Data source triangulation can help address self-reporting (Denzin 1984; Stake 1995), drawing data from multiple sources, pre-, post- and even during coaching and not just from the coach and the coachee but also from other individuals in the coachee's environment.

Whether –and how- the social context is considered in existing EC outcome studies

EC is a leadership development intervention. As such, it is context sensitive (Martineau & Patterson, 2010, p. 274). It is also a social process. Orenstein (2002, pp. 355 and 364) noted that individuals -including the coach and the coachee- are influencing and being influenced by their interaction and the multi-dimensional system that the organization represents. As such, we count the individual actors and the intervention as key social context ingredients. We believe that the *social context* of an intervention must be actively considered in outcome studies, hence, we set out to explore how it has been treated in the EC research so far.

Leadership is "a social and goal-oriented influence process, unfolding in a temporal and spatial milieu" (Fischer et al., 2017). So is EC. Yet, most EC outcome studies overlook the social context, with more than half (60 of 110) having no consideration of the interrelation between EC outcomes and the relational, spatial or temporal context within which they take place. Specifically, we found that outcome studies belong to three categories (Table 1): a. 50 studies adopt a *sociological approach* [e.g. Grant's (2014) study of outcomes during organizational change or De Haan et al.'s (2013) and Smith and Brummel's (2013) coaching "active ingredients" research], b. 33 studies take a *descriptive* approach where social context is acknowledged but not used in the analysis [e.g. in Styhre's (2008) study the construction industry is a backdrop rather than an active part of the data analysis or Anderson's (2002) and Gray et al.'s (2014) studies that vividly describe coachees' intervention experiences but hardly refer to influences from the environment within which these happen] and c. 27 "*acontextual*" studies -a term borrowed from Pettigrew (1985)- where context is absent. Interestingly, we found a notable increase in recent years of studies that take a contextual approach with 13 of the 50 context-sensitive studies published between 2015 and 2016.

Do impact factor journals publish EC studies with better research designs?

As aforementioned, one of our screening methods was to separate the impact factor studies from the non-impact factor peer-reviewed studies. Before proceeding to this second part of our research that focuses on EC outcomes and their contextual influences, we tested whether these two subsets really differed in terms of the quality of research designs. From the 110 peer-reviewed studies, 78 are published in journals that do not have an impact factor and 32 have. We found in the impact factor studies, too, some design weaknesses, but proportionately less in comparison to the non-impact factor ones. Only one of the 32 (i.e. 3.1%) impact factor studies (Orenstein, 2002) uses a narrative approach consisting of case history/vignettes of coachees, compared to 13 of the 78 non-impact factor peer-reviewed studies (16.7%). Only in two of the 32 impact factor studies the author is the coach (6.3%) compared to 22 of the 78 non-impact factor ones (28.2%). 14 of the 32 impact factor studies (43.75%) use a research design that involves collecting data at different points in time even if that is simply pre/post-coaching, compared to 16 of the 78 non-impact factor peer-reviewed studies (20.5%). Somewhat a higher proportion of impact factor studies (5 of 32; 15.6%) use experimental or quasi-experimental designs compared to non-impact factor ones (10 of 78; 12.8%). Last but not least, none of the impact factor studies focus on ROI evaluations.

Key challenges for EC research

Irrespective of research methods employed, there are several methodological challenges that the field needs to address. In their systematic review, Grover and Furnham (2016, p. 26) noted that EC must be mindful of emerging concerns from the psychology literature after a recent research showed that more than half of a sample of 100 published psychology studies were not replicable even when researchers used the original materials. In addition to the research dependence issue that we identified in a substantial subset of EC studies, we consequently discuss based on our review what we see as some key EC research challenges:

Short-term outcomes

Day (2000, p. 586) observed how leadership development is a continuous process that happens beyond the "classroom" as training location and is inextricably linked to one's work context. The same observation applies to EC. It is in the work context that the intervention outcomes are applied and tested. Day-to-day pressure may prove a crucial influence on whether the coachee can make sustainable changes. Since the intervention is not independent of its social context, the problem of self-reporting of outcomes is often linked to the challenge of evaluating the sustainability of the intervention. Pre/post-coaching designs, ex-post surveys or qualitative interviews at a single point in time post-coaching, without consistent measurement of changes in the long-term, offer a distorted version of reality and assume sustainability of outcomes. As Fischer, Dietz and Antonakis (2017, p. 1735) note on research on leadership processes, "such designs make implicit temporal assumptions regarding the immediacy of effects and/or a stable equilibrium among these effects". With regard to leadership development initiatives, Packard and Jones (2015) found that by obtaining supervisory ratings on participants' job performance and participants' own performance rating, it was possible to moderate the problem of providing data from self-perceptions rather than actual behavior. The use of multiple stakeholders' ratings on coachee's behavior is a way to address the self-reporting challenges associated with EC outcome research, too.

Research designs need to be longitudinal to allow tracking changes in the months or years following the intervention to evaluate sustainability. Fischer et al. (2017) propose the use of an event-based "person-parts" approach, whereby events are seen as "episodes where actions intersect with a context" (p. 1736). This approach is promising and reflects recent EC outcome studies on the critical incidents of EC interventions allowing to capture the social context and actors' interaction at several points in time (De Haan, Bertie, Day & Sills, 2010; De Haan & Nieß, 2012; Turner & McCarthy, 2015; Diochon & Nizet, 2015).

Common method and endogeneity bias

Besides self-reporting, there are several *method biases* observed in many EC outcome studies. Here we define and discuss the issue of *endogeneity*, including more specifically the issues of *common method bias* and *omitted selection*.

Endogeneity refers to the incorrect modelling of causal relations which makes it impossible to interpret the effect of one variable on another (Antonakis, Bendahan, Jacquart and Lalive, 2010). Two

examples are the correlation of disturbances of potentially endogenous regressors in mediation models without testing for endogeneity or comparing a treatment group to non-equivalent groups (Antonakis et al., 2010). *Method bias* is widely perceived as the use of uncontrolled method factors in a study, hence causing bias to "estimates of construct reliability and validity" or to "parameter estimates of the relationship between two different constructs" (Podsakoff, MacKenzie & Podsakoff, 2012, p. 542). Siemsen, Roth and Oliveira (2010, p. 457) define common method bias as "the degree that estimators become inconsistent; i.e., parameter estimates asymptotically converge to values different from their true population value". On the other hand, *omitted selection* is observed when "comparing a treatment group to other non-equivalent groups", "comparing entities that are grouped nominally where selection to group is endogenous" or having a sample that "suffers from self-selection or is non-representative" (Antonakis et al., 2010, p. 1091).

Antonakis, Bendahan, Jacquart and Lalive (2010) reviewed 110 articles on leadership in top-tier journals and observed a staggering range between 66% and up to 90% of design and estimation conditions that the researchers failed to address, meaning that the causal claims these studies made were, in fact, invalid. This was attributed to *endogeneity* manifested in any of 14 validity threats. Similarly, Theeboom et al.'s (2014, p. 14) statistical meta-analysis showed a lack of rigorous study of the causal mechanisms by which coaching interventions are effective. For instance, certain coachee behaviors, such as self-efficacy, are treated in some quantitative studies as an outcome variable measuring coaching effectiveness, whereas in others as an individual difference or trait predicting coaching effectiveness (Grover & Furnham, 2016, p. 24).

Such weaknesses, often the result of poor research designs, can be addressed by drawing on solutions from economics and behavioral research (e.g, Angrist and Pischke, 2009; Podsakoff, MacKenzie & Podsakoff, 2012). For instance, as far as *method bias* is concerned, Podsakoff et al. (2012) suggest several procedural remedies such as obtaining measures from different sources; application of temporal separation (i.e. time delay), proximal separation (i.e. physical distance) and psychological separation between predictor and criterion as well as statistical remedies. An example of proximal separation is the collection of EC data outside the physical context of the intervention (e.g. the sponsoring organization premises). An example of psychological separation is the embedding of EC-related questions within a broader set of non-EC questions on leadership development or work-life balance to make it less evident to participants that the collected data are specifically about EC.

Among the different types of biases and validity threats, we find "omitted selection" (Antonakis et al., 2010, p. 1091) to be most concerning as it cuts across qualitative and quantitative studies. The self-selection of EC participants is an example of this. While some studies report resistance or cynicism from coachees (e.g. Laske, 2004; Gray, Gabriel & Goregaokar, 2015), most participants are coached because they want to (hence, are positively inclined to the intervention) and are likely to expect positive outcomes. How can EC findings be protected from omitted selection? For instance, De Haan et al.'s (2011) found that the coaching relationship and positive expectations contribute to positive EC outcomes. It would be difficult to know if the positive outcome is the result of good EC or of the individual's positive expectation of it.

Antonakis et al.'s (2010) suggested various research strategies to address potential endogeneity. These could apply in EC research, too, such as: "fixed-effects panel" (i.e. collecting repeated observations of coachees to capture comparable hierarchical or longitudinal data across cases), careful "sample selection" for the treatment and the control groups and difference-in-differences models (i.e. "compare a group who received an exogenous treatment to a similar control group over time"; Antonakis et al., 2010, p. 1099).

Though these strategies are drawn from -and typically are applied to- quantitative research, the logic behind them, such as careful sampling and collection of data at different points in time, also apply to qualitative research.

Demand effects induced by authors: Hawthorne and social desirability effects

Since personality measures are vulnerable to response distortion due to "social desirability" influences (Edwards, 1990; Ellingson, Sackett & Hough, 1999; Hogan & Nicholson, 1988; Nicholson & Hogan, 1990), the construct validity of outcome studies can be challenged. Experimental research in economics (Zizzo, 2010) has showed that participants may receive either social or purely cognitive cues (i.e. participants' understanding of the experiment) about what constitutes appropriate behavior and change behavior accordingly. Social desirability influences are observed in all sorts of methods, not just experiments. Social cues are particularly strong when the coach is the author and has established long-term rapport with participants. Another problem is the "methodological Hawthorne effect" first observed in experiments by Adair (1984). Subjects are likely to change behavior compared to how they would have behaved had they not known about the research (see Sturm and Antonakis (2015) on the pitfalls of priming participants in experimental designs). Similarly, when participants are selected by their organization to be coached -a leadership development intervention now offered by companies to high-potential executives (Coutu et al., 2009)- they are more likely to report positive outcomes. They would want to signal to their coach and their sponsoring organization that they make progress. Data source triangulation (Denzin, 1984; Stake, 1995) can help overcome such biases by drawing data from multiple sources such as surveys of different types of coach-coachee dyads (e.g. Sonesh et al., 2015) or self-reporting combined with 360-degree feedback (e.g. Howard, 2015).

Unfair comparisons of interventions

Some studies appear to make unfair comparisons between interventions, constructs or variables that may not be comparable. The use of different outcome measures (e.g. self-awareness, goal attainment or other individual-level outcomes) has led to large variability of effect sizes due to these being distinctively different constructs (Grover & Furnham, 2016). Similarly, research that compares EC with other developmental interventions (e.g. mentoring; Salter, 2014) may not lead to conclusive results as to the comparative value of the two since each is a different practice. When two or more theories are tested or when two or more variables are compared for their predictive or relational strength, one may prove stronger than the other or others due to it being "more strongly operationalized, manipulated, or measured" and hence, favored (Cooper & Richardson, 1986, p. 179). One could then question whether in an experimental EC outcome study (e.g. Grant, Green & Rynsaardt, 2010; Moen & Federici, 2012a, 2012b; Ladegard & Gjerde, 2014), the comparison between a coached group and a non-coached, control group is the best research strategy. Sturm and Antonakis (2015) find such a comparison potentially problematic since participants may tend to behave in a certain way because of knowing the topic of the study (i.e. social desirability issues). Podsakoff et al.'s (2012) proposed procedural remedies -that allow the separation of predictor and criterion measures- may provide some help in this. For instance, as mentioned earlier, one remedy is psychological separation where the data collection instruments are designed in such a way that it is not apparent to participants that the study is EC-specific (e.g. by including non-EC items). Another remedy is the proximal separation of the coached and the control group so that the one group is unaware of the other and vice versa. Hence, neither can be influenced by the knowledge of a possible comparison with one another. Therefore, an EC study design may influence the effects of certain outcomes more than others (Grover and Furnham, 2016, p. 7), which would explain why Theeboom et al.'s (2014) meta-analysis showed that a study's research design has an effect on EC outcomes whereas Jones et al.'s (2016) showed that it has not. What would account for a clean counterfactual

to a coaching group can be both theoretically and practically challenging when designing an EC outcome study.

Cooper and Richardson (1986) suggest that unfair comparisons may be the result of poorly framed empirical questions with the study outcomes being at least partial artifacts of the procedures, manipulations or measures used in the design. One of the solutions that Cooper and Richardson (1986) suggest involves having "independent judges" (p. 182) to read the descriptions of theories, factors or variables. For instance, in a qualitative study on the critical moments in an EC intervention, EC practitioners and coachees could be invited to read descriptions of various events accounted as "critical moments". The purpose would be to cross-check for clarity in their definition and understanding. Another solution proposed by Cooper and Richardson (1986) in order to address the issue of unfair comparisons is to select sample with adequate variance on the variables under investigation. For instance, this could mean selecting an EC study sample that is gender-balanced. Moreover, a further solution that is advisable in experimental studies is that researchers conduct "manipulation checks" as a "precaution" prior to proceeding with the data collection to establish whether the manipulations worked (Cooper & Richardson, 1986, p. 183). Last but not least, researchers' honesty about any procedural or distributional inequalities is also key (Cooper & Richardson, 1986).

Other challenges that are specific to qualitative research

While research endogeneity is perceived as predominantly a peril of quantitative research, some of the challenges discussed such as self-reporting, drawing on short-term outcomes, Hawthorne and social desirability effects and the unfair comparison of interventions are observed in both qualitative and quantitative studies. According to Lee (1999; in Pettigrew, 2013, p. 124) while "the quantitative researcher may lean towards prevalence, generalizability, and calibration, the qualitative researcher gives greater emphasis to narration, description, interpretation, and explanation"; the "best qualitative work is contextually grounded and seeks to understand process dynamics and not just outcomes". As such, qualitative methods are particularly suitable for the *how* and *why* research questions of organizational phenomena, particularly when change is involved (Gray, Stensaker & Jansen, 2012). This renders qualitative research quite fitting to the study of EC as a personal development intervention that encompasses change. The problem with qualitative research is, though, a lack of consensus regarding the evaluation criteria used (Pratt, 2008). The non-standardized data collection and analysis tools of qualitative research (Bluhm, Harman, Lee & Mitchell, 2011), not having a "boilerplate" and not agreeing on whether such boilerplates are "acceptable" make publishing qualitative research more challenging (Pratt, 2008, p. 495). A further shortcoming is its lack of statistical generalizability, which is compensated by analytical generalizability whereby data are generalized based on a theory, not on a sample (Pratt, 2008; Yin, 1994). In qualitative studies construct validity, external validity and internal validity tests are seen as ways to ensure robust research designs (Yin, 1994, pp. 32–38). However, with a lack of consensus about qualitative research evaluation, even these measures do not ensure that a produced research will be unanimously approved as robust. Pratt (2008) surveyed qualitative researchers who have published in top-tier North American journals and found that a common complaint among qualitative researchers was that their research was "often, and inappropriately … judged based on criteria more appropriate for quantitative papers, or perhaps more accurately, positivistic, and deductive research". This is a challenge we faced with regard to the evaluation of EC research, as we tried to appreciate each method – quantitative or qualitative- for what it is, while remaining critical of any design shortcomings in the reviewed studies.

To make qualitative research more robust, one could collect interview data, archival data and observation using best-practices in qualitative research such as longitudinal designs, accuracy checks, triangulation, and in-depth analysis (Bluhm, Harman, Lee and Mitchell, 2011). Pratt (2008) even suggests that qualitative research mimicks quantitative research regarding its structure and evaluation criteria and that qualitative data are combined with quantitative data.

Finally, researcher bias and the reduction of data to produce meaning are generally accepted characteristics of qualitative research (Bluhm, Harman, Lee & Mitchell, 2011). For the former, both Pratt (2008) and Bluhm, Harman, Lee and Mitchell (2011) stress the importance of transparency in the methods and analysis. Aa far as the latter is concerned, Klag and Langley (2013, p. 162) note that the "detailed mechanics of conceptual leaps remain as elusive in management research as they do in the world of art". However, a careful consideration of the theoretical purpose of the qualitative study from its outset (Bluhm, Harman, Lee & Mitchell, 2011) helps to set the path for a better crafted theoretical contribution.

What we know and do not know about coaching outcomes (Q2a)

Bridging objectives 1a and 2a: screening studies in preparation for question 2

Since our research question 2 is about what we know regarding the outcomes of EC and in the previous section we identified several weaknesses in current EC outcome study designs, we decided to further screen the studies on which we will base our discussion on EC outcomes. Here, we should note that with this being a qualitative, systematic review we did not seek to compare the strength of the different outcomes or contextual factors identified in the literature, but instead present what current research has found. What we did was to order the research designs in terms of their robustness and confidence they offer with respect to research validity. We screened out only those studies judged as highly unreliable.

Despite some of the methodological weaknesses noted earlier, overall, RCSs or quasi-experimental studies are robust designs since they allow testing differences in outcomes between control and coached groups. Well-designed qualitative studies, such as case studies (e.g. Gray et al., 2011a; Baron & Morin, 2009), are useful too since they allow answering context-sensitive "why" and "how" questions. Overall, we regard good quality studies those that test or enrich their findings using multiple data sources (e.g. Kombarakaran, Yang, Baker & Fernandes, 2008) or multiple research methods (e.g. Gray et al., 2011a), including data from multicourse feedback and other assessment tools (e.g. Wasylyshyn, Gronsky & Haas, 2006) as well as studies with repeated measure designs, pre-/post- and even during coaching (e.g. Jones, Rafferty & Griffin, 2006). There are single method studies -whether interview (e.g. Toogood, 2012; Cowan, 2013) or survey-based (e.g. Hooijberg & Lane, 2009)- and often only from one type of stakeholder (e.g. Toogood's (2012) in-depth interviews of coaches). Less robust are the single method studies that rely on coachees' self-reporting without pre −/post-coaching data (e.g. Stevens, 2005) as well as case narratives or histories of coachees without cross-case comparisons (e.g. Diedrich, 1996). At the bottom of the hierarchy are the two ROI and the 24 author-as-coach studies (impact and non-impact factor ones) which we excluded for the purpose of answering question 2; however well-designed some of them were. This produced a dataset of 84 studies. We debated the inclusion of single method and coachee self-reporting studies but decided to keep them because unlike coaches, coachees would not normally have control of the published material and normally have less self-interest in reporting positive EC outcomes compared to coaches' self-reporting.

What we know: evidence from the EC outcome studies

In Table 2 we list the outcomes identified in the 84 screened studies. We found more than 70 positive outcomes (including examples in the second column of Table 2). We grouped these outcomes into 11 categories (first column, Table 2). We came across only 16 studies that report either negative or not-fully positive outcomes which we would

describe as moderate or conflicting and grouped them into eight types of EC pitfalls with some observed in more than one study. These are presented in Table 2 (last column) as questions or challenges practitioners must keep in mind. As we have already screened out studies with potential outcome biases, such as the author-as-coach ones, this smaller dataset is less likely to include biased positive results. We also checked whether the 15 experimental/quasi-experimental/RCS studies, which are at the top of the design robustness order, report more often negative outcomes. Only three experimental/quasi-experimental studies report negative or moderate (i.e. not clearly positive, often unintended) outcomes. Smither et al.'s (2003) study reported a moderate finding where managers who worked with an EC improved more than other managers in terms of direct report and supervisor ratings, but the effect size was small. Bozer et al. (2013) observed a negative relationship between learning goal orientation and improvement in self-reported job performance among coachees with low levels of pre-training motivation. Finally, Bozer, Baek-Kyoo & Santora's (2015) pre-/post-test design showed that male executives with female coaches reported that their self-awareness was not improved by EC. The experimental study findings also confirm each of the positive outcome categories from 1 to 10 in the first column of Table 2.

The positive outcomes in Table 2 are grouped per stakeholder. Laske (2004) identified two types of EC positive effects: “behavioral” and “developmental”. We found three types of positive outcomes for the coachee: personal development, behavioral changes towards others and work performance. An example of a positive *personal development* outcome is that coaching helps coachees improve their resilience, workplace well-being and reduce stress (Grant et al., 2009, RCS). Another finding is how EC improves work performance and planning (Hall et al., 1999, interviews; Bowles, Cunningham, De La Rosa & Picano, 2007; repeated measures design; Fischer & Beimers, 2009, mixed methods; Moen & Skaalvik, 2009, Moen & Federici, 2012a and Bozer et al., 2013, experimental design).

Without any ROI studies in our dataset, we looked for other organizational-level outcomes. Nine studies identified indirect, positive organizational-level effects. EC appears to impact on organizational performance in non-ROI terms [e.g. Gorringe's (2011) case study of improved NHS waiting times; see also Luthans and Peterson, 2003; mixed methods; Levenson, 2009; mixed sources] and helps deal with organizational change (Grant et al., 2009; RCS).

The potential effects on coach's personal development (Hall et al., 1999, mixed methods; Toogood, 2012, interviews) are the only category of positive outcomes not yet confirmed by RCSs. They are interesting findings, though, that warrant further investigation.

Although EC outcome studies consistently report positive outcomes, a few did identify areas where EC may produce conflicting outcomes (mostly either unintended or less positive than expected) or negative. These constitute potential EC pitfalls (Table 2, last column) which we organize around the various coaching phases:

a. *Decision to engage in coaching* (Pitfalls 3, 5 and 8, Table 2): With the exception of Smither et al.'s (2003) quasi-experimental study, showing very small positive effect on coachees' behavior compared to non-coached participants, all other experimental studies found significantly more positive outcomes for coachees. Despite its potential, EC requires organizational and individual buy-in considering the resources and time investment required. For instance, Polsfuss and Ardichvili's (2008) interview-study showed that some managers do not see the merit or value of certain coaching practices. EC also requires from coachees a positive attitude and openness to change. However, this is not always the case (e.g. Gray et al.'s (2015) study of unemployed coachees found some being cynical towards coaching). The EC engagement decision also involves choosing among different forms of EC and settings, which have varying effect on EC processes and outcomes. For instance, EC within an executive education context is of shorter term and coachees have different needs and expectations from their coach compared to other coaching (Hooijberg & Lane, 2009, survey).
b. *Contracting and designing an intervention* (Pitfalls 1, 2, 7; Table 2): A potential pitfall in the contracting and design of an EC intervention relates to coachee selection. This is a context-sensitive decision based on an organization's leadership development and succession management plans, which, however, begs the question whether EC is right for everyone. Research has showed that EC purchasers define coaching success (i.e. positive business results) differently from coachees (i.e. personal development benefits) (Leedham, 2005, mixed methods/sources). The challenge is to identify outcomes that meet stakeholders' needs and expectations and to regularly measure them (Ely et al., 2010, p. 588). This is as important for clients as for coaches. Unfortunately, in practice only few coaches measure the outcomes of their work (Grant & Zackon, 2004, survey).
c. *Implementation* (Pitfalls 4 and 6, Table 2): Coachees' goal orientation and commitment contributes to positive outcomes, but when they set too difficult goals, devote less time to accomplish them or have low pre-coaching motivation then coaching-induced goal-setting can negatively impact on job performance (Bozer et al., 2013, non-randomized controlled study; Bowles & Picano, 2006, survey). This means that EC goals must be *pragmatic* (e.g. not too difficult to achieve, having time availability to work towards them) and must also be *relevant* (e.g. goals matching coachee's motivation) with leeway for possible adjustments as the intervention progresses. Finally, they should also *fit the intervention context*. For instance, Gray et al. (2011a) found that in small and medium sized enteprises coaching is used mostly as a personal, therapeutic intervention than for business-oriented competency improvement.

Single-method data collection, self-reporting and ROI: reflections from the data

As mentioned earlier, we debated whether to include single method and coachee self-reporting studies but decided to include them. We also debated whether to include author-as-coach studies and ROI studies and decided *not* to include them. We explained why we made such decisions. Here it is worth discussing how different Table 2 would have been if the former (single method and coachee self-reporting) had not been included and also how different Table 2 would have been if the latter (author-as-coach and ROI studies) had been included. We will briefly reflect on the research questions and related outcomes these studies investigate, as well as the quality of the research designs employed.

As far as the single-method and the coachee self-reporting studies are concerned, what we observe is that when the single method is a survey then most of studies would focus on either the coachee or examine aspects of the coaching practice and to less extent focus on outcomes relating to the coach. On the other hand, the study of EC outcomes relating to the coachee or to the coach are the key areas of focus for most studies using interviews. The fact that survey-based studies are comparatively less focused on the coach and interview-based studies are less focused on the coaching practice may be attributed to access issues (as a population, are coaches perhaps more accessible for in-depth qualitative interviews?). On the other hand, it is surprising that surveys seem to be preferred for the study of research questions on coaching practice, since “how”/practice questions would have been better tackled with a qualitative research design. For instance, De Haan, Duckworth, Birch and Jones' (2013) survey-based study explores the “active ingredients” that predict the effectiveness of EC and determine the difference in predictive value of these ingredients on EC effectiveness. Despite being a very well-crafted study, the chosen method reflects a limitation in single method studies. The survey offers insights into the list of active ingredients, but the nature and effect of these ingredients can be studied in more depth with qualitative methods. This re-iterates our focus on the importance of using mixed

methods in EC outcome research.

As far as coachee self-reporting is concerned, most of these studies focus as expected on research questions about coachee outcomes and some also focus on aspects of the coaching practice (e.g. Gan & Chong's (2015) study on the nature of the coaching relationship and its link to coaching effectiveness). Some coachee self-reporting studies are quantitative (e.g. Moen & Allgood's (2009) study on the impact of EC on leadership self-efficacy) and others qualitative (e.g. Stevens' (2005) study drawing on informal telephone interviews with seven managers). While we did not observe any significant design flaws among the single method studies and the coachee self-reporting studies, we conclude that the use of mixed methods would make their findings more insightful and the design more robust. Had we not included the single-method and self-reporting studies in Table 2, the only substantial difference we would have had is somewhat less detail regarding the variety of EC outcomes presented in the Table 2 - particularly those outcomes relating to coachees.

Finally, we reflect on how the inclusion in Table 2 of the research dependent (coach-as-author) and ROI studies would have changed this Table. Firstly, in only one of the research dependent studies are any negative outcomes reported (lack of developmental advance among certain participants and resistance to change during the coaching period; Laske, 2004). Impressively, nearly all of the research dependent studies (22 of the 24) tend to prefer a case study approach. One of the remaining research dependent studies was survey-based (Wasylyshyn, 2003) and one used mixed methods (Tamir & Finfer, 2016).

For 14 of the 22 case study-based research dependent studies, the focus is individual coachees. Typically, without any robust research design and with lack of triangulation, these studies present narrative accounts -namely, the coach-as-author's account- of how one or more coachees improved as a result of the EC they offered (e.g. Diedrich, 1996; Winum, 2005). The remaining eight case studies are organization-focused and discuss the outcomes from several executives of the organization having been coached by the author. The case studies of organizations are somewhat better designed with data collected either quantitatively (e.g. Wales, 2003), from multiple sources (Gaskell, Logan & Nicholls (2012) or with repeated measures (Lawrence, 2015). However, overall, with few exceptions, they tend to be of rather limited methodological rigor. Despite the case studies being perceived as a method more likely to be attuned to social contextual considerations, 14 of the 24 research dependent studies offer a descriptive account of context, one is acontextual and the remaining nine take an active consideration of context. The latter would refer to culture, motivations, politics and industry-specific influences on EC. In terms of focus, nearly half (11) of the 24 research dependent studies have research questions relating to the coachee (e.g. Freedman & Perry, 2010) and less focus on coaching as an intervention (i.e. not focusing on stakeholders per se) (e.g. Wasylyshyn's (2003) study on different aspects and success factors associated with EC). In summary, while the research questions these research dependent studies seek to address are interesting and relevant, their research design and execution are not as robust. This also applies to ROI studies as discussed earlier. A possible explanation, particularly with regard to the research dependent studies, is that often the authors are practitioners without formal research training, hence more likely to carry out poorly designed and executed research.

Drivers of coaching outcomes as identified in existing outcome studies

So far, we established *what we know* about EC outcomes – positive outcomes and potential pitfalls. Now we turn our attention to reported factors that can influence these outcomes. Such a review will help identify what we *still do not know* about EC. We found a limited but growing number of studies that look into such factors (e.g. Sammut, 2014; De Haan et al., 2013; Smith & Brummel, 2013; De Haan and Nieß, 2012; De Haan et al., 2010). We present them in Table 3, organized into five categories relating to: a. *coaching intervention/process* (e.g. critical moments of coaching), b. *organization* (e.g. organizational support), c. *coachee* (e.g. personality and personal attributes), d. *coach* (e.g. coach's background), e. *relationship of stakeholders* (e.g. stakeholder alignment on coaching purpose/goals). We checked which factors have been identified in experimental studies. Only a few have and are marked on Table 3 with an asterisk (*). In the second column of Table 3 we summarize the evidence on how these contextual factors contribute to positive outcomes. Some of these factors are more direct (e.g. long-term coaching is found to be more effective than short-term; Thach, 2002), whereas others are less overt (e.g. coach's use of influence tactics during the intervention increases coachee's commitment; Lewis-Duarte & Bligh, 2012).

The contextual factors (column 1; Table 3) and how their influence is manifested (column 2; Table 3) could serve as roadmaps for practitioners to fine-tune an intervention as it progresses, based on coachee's and the organization's response to it. Such fine-tuning could be as subtle as mimicking the client's language (see “language and communication” under the “Intervention” category; Table 3) or as prescribed and focused as including alignment on purpose, methods, measurement criteria and desired outcomes in the contracting of the engagement (see “stakeholder alignment category”; Table 3). We also observe from the mapping of contextual influences that each type of outcome -whether it relates to individual stakeholders, to the group or the organizational level- does not happen in isolation. The meshing of contextual factors reflects the meshing of contextual levels of EC outcome analysis. For instance, organizational support (under the Organization category; Table 3) as a contextual influence has effects on a coachee's expectations and sensemaking of the intervention (under the Coachee category; Table 3) and this, has effects on EC outcomes.

We hope this list of factors can guide practitioners to areas they need to focus on –and where possible manage. These factors also offer researchers an overview of the range of influences on EC, but also the data on Table 3 trigger the need for research that explores further the interconnectedness of EC outcomes and the contextual factors that shape them.

Implications for research and practice

We now proceed to the second part of our research questions 1 (objective 1b: strengths and weaknesses in the use of different research methods in future EC outcome studies and design implications) and 2 (objective 2b: implications for practice and research regarding the evidence on EC outcomes). In summary, we have so far provided an overview and evaluation of how EC outcome studies are designed and what their focus is. Overall, a growing number of EC outcome studies are produced each year, which is a promising development. Among the strengths of these studies is the pluralism of themes that allow exploration of multiple aspects of EC. The field tends to become increasingly aware of the importance of choosing appropriate research methods to increase the reliability of reported outcomes. However, we found several weaknesses particularly regarding the research designs concerned. Most research focuses on the coachee, much less on the coach and even less on the sponsoring organizations. There is a dearth of studies that compare coaching methods to test which are more effective or whether other factors are more important for coaching success. Also, a considerable number of studies rely on self-reporting either by the coachee and/or the coach. We expressed doubts on the value and validity of ROI as an effective measure of EC outcomes. Last but not least, we found that social contextual considerations are not well-addressed in current EC outcome studies despite the role of context on any EC intervention. We identified the range of outcomes that EC reportedly produces and how these are spread across coaching stakeholders (Table 2). We also identified the range of contextual influences on EC outcomes and how these are manifested in existing outcome studies (Table 3). We urge the field to use these findings as the basis for the design of better research and practice.

Table 4
Research methods selection for future executive coaching outcome studies: strengths, weaknesses & other issues.

Research method	Strengths	Weaknesses	Design issues/variables that need to be included	Other methodological considerations
Case study	EMPHASIS: Social context of intervention; the coaching experience (e.g. critical moments, process and method) BREADTH VS. DEPTH: Depth	Transferability of findings to different contexts Less emphasis on the "what" questions (content of coaching)	Data from all 3 stakeholders (coach, coachee, organization) Combine qualitative data with personality assessment tools, including multisource feedback	Collect longitudinal data Ensure the coach is not the author Ensure skillful interpretation of data
Survey/quantitative (only)	EMPHASIS: Attitudes towards coaching or perceptions of effectiveness (e.g. coachee's satisfaction); EC evaluation issues BREADTH VS. DEPTH: Breadth	Little emphasis on social context of intervention Less emphasis on "how" and "why" (i.e. process, feelings, experience)	Focus on one type of stakeholder (e.g. coachees' perceptions) or multiple stakeholders Draw also data on personality assessment tools, including multisource feedback	Craft longitudinal/repeated measures design Report statistical power based on N Ensure the coach is not the author
Qualitative interviews (only)	EMPHASIS: Perceptions of effectiveness of coaching; experience and/or thoughts on coach intervention BREADTH VS DEPTH: Both	Less in-depth study of context than case studies; typically smaller data pool than surveys Less emphasis on the "what" questions (content of coaching)	Focus on one type of stakeholder (e.g. coachees' perceptions) or multiple stakeholders Draw also data from personality assessment tools, including multisource feedback	Craft longitudinal/repeated measures design Ensure the coach is not the author Ensure the sample size allows identifying emerging patterns
Experimental/quasi-exp./ randomized controlled studies	EMPHASIS: Testing specific variables (e.g. self-efficacy, need satisfaction, goal-setting behavior, motivation); comparison with control group BREADTH VS. DEPTH: Depth	Complexity of research design Difficulty in obtaining data access Research costs	Focus typically on coachee (no RCSs studies yet of coaches) Design tests specific developmental, behavioral or performance variables Multiple methods and sources (survey, interviews, personality assessment tools, including multisource feedback)	Craft longitudinal/repeated measures design Careful sampling and comparison of control versus coached group Ensure skilful interpretation Ensure the coach is not the author
ROI	EMPHASIS: Link between individual and organizational outcomes BREADTH VS. DEPTH: Depth	Relies mostly on estimates or self-reporting Doubts about value of ROI calculations	Instead of financial ROI focus on other measures (e.g. client satisfaction) Research design based on multiple data	Collect longitudinal data Decide on trade-off of intangible vs. tangible benefits; who decides what to measure Ensure the coach is not the author
Meta-analysis/systematic reviews	EMPHASIS: Field-level trends and evidence of outcomes BREADTH VS. DEPTH: Both	Challenge of comparing studies of varied research quality and methods	Content analysis Systematic review	Determine research quality Identify parameters to study

Strengths and weaknesses of research methods and research design implications (1b)

Grover and Furnham (2016, p. 26) observed that "as an industry, coaching needs more stringent methodology, statistical analysis and larger sample sizes to increase the generalizability of the coaching effectiveness findings". We agree on the need for a more stringent methodology but take a quite different tack about the way to do so. We argue that an emphasis on statistical analysis and large sample sizes is not the only way forward for EC outcome research. Table 4 summarises the strengths and the weaknesses of different research methods.

We believe that favouring one method over another is limiting. It is more important to craft good research questions and identify the best research design that fits the questions, rather than favour one method over another. Some of these designs seek research depth and have a more context-sensitive focus on micro-themes or look into the interrelationships between levels of EC outcomes where case studies, qualitative interviews or RCSs/experimental designs would be more suitable. Other designs seek breadth rather than depth or a balance between the two and would render surveys or meta-analyses as more suitable methods. We encourage the mixing of data collection methods (e.g. surveys with interviews). Indeed, as we have observed, multimethod and multi-source approaches help to overcome several of the research design shortcomings we identified in current studies. Bluhm, Harman, Lee and Mitchell (2011) reviewed qualitative research published between 1999 and 2008 in several 'major' US and European management journals and found that the more data collection methods a study used, hence offering increased validity from triangulation, the larger the influence of the article on the accumulation of management knowledge, meaning that the study was more frequently cited. Next, we reflect on the methods presented in Table 4:

Case study research enables the exploration of the *social context* and *experience* of EC (e.g. De Haan and Nieß' (2012) study of critical moments). Its primary weakness is the low transferability of findings across contexts. A thorough case study design draws data from multiple stakeholders and combines qualitative data with multisource feedback or other tools. Case studies and experimental studies are most promising in providing insights into contextual influences or individual-level moderators in coaching practice. In case study research, robust designs can be developed by using at least one of three validity tests. It involves using multiple case studies (construct validity), replication of data findings (external validity) and identification of patterns and causal relationships (internal validity) (Yin, 1994 pp. 32–38). Eisenhardt & Graebner (2007) note that there is a trade-off between presenting a rich story from the empirical evidence and contributing with well-grounded theory.

As noted earlier, despite several calls by leadership scholars (Day, 2014; Shamir, 2011; Day & Lord, 1988; in Fischer et al., 2017) the field's conceptual and empirical work has neglected the role of time and needs to study "how effects unfold and last over time" (Fischer et al., 2017, p. 1735). Following on Denis, Langley and Sergi (2012), Fischer et al. (2017, p. 1742) describe the study of processes in natural settings -such as case studies (Yin, 1994)- as an underexplored avenue that could help better understand the multi-path nature of leadership effects and help build theory on process-driven leadership mechanisms. EC, a context-sensitive leadership development practice, fits perfectly in that category. We identified 26 EC case studies of organizations and 18 case studies of individuals; some better than others. Fischer et al. (2017) also highlight the use of other contextually rich sources of data such as historiometric data (Simonton, 2003). Such data have not yet been employed in EC research, but with new studies drawing on textual analysis of the transcripts of coaching sessions (e.g. Kauffman & Hodgetts, 2016), this may be a method to consider in future research.

Survey-only EC outcome studies focus on attitudes and perceptions of effectiveness and EC evaluation issues. Surveys tend to ignore the social context of the intervention, unlike qualitative interview studies, which tend to have a smaller sample but are more in-depth. Reflecting the field's need for more longitudinal research, surveys at repeated points in time can prove a particularly powerful method to track trends on EC outcomes. For each research design included in Table 4, we discuss, besides their strengths and weaknesses, also what design issues researchers need to account for when applied in the study of EC outcomes. For instance, when designing a survey, researchers need to determine whether the focus will be on one type of EC stakeholder (e.g. coachee's perceptions of EC effectiveness) or whether the survey design will compare perceptions across different stakeholders (e.g. coachees, coaches, HR professionals). Methodological considerations for the surveys would include whether the survey would run at different points in time to capture longitudinal trends, what the sample size should be and whether the selected sample would be independent of the author or not (i.e. coach-as-author implications). Studies based on semi-structured qualitative interviews may require similar considerations as surveys (e.g. sample size but of a different scale, research independence and a longitudinal design to track changes over time).

Scholars have called for coaching outcome studies that use control groups, randomized experiments and rigorous statistical analysis (De Haan & Duckworth, 2013; Ely et al., 2010). Such studies (e.g. Ladegard & Gjerde, 2014; Moen & Federici, 2012b; Grant et al., 2010) provide robust proof that coaching is effective (De Haan & Duckworth, 2013) and represent a scientific, evidence-based research approach from which the field can greatly benefit (Griffiths & Campbell, 2008). They are typically designed to test impact in relation to specific variables such as need satisfaction or self-efficacy. They may have room for contextual considerations, but less so than case studies do. The main challenges with experimental studies are difficulties in research access, the implementation of the design and interpretation of findings and the fact that it is often a costly method. This perhaps explains the few (15) experimental/quasi-experimental/RCS EC studies so far. Other issues that need to be considered in the design of RCSs are the type of variables tested, whether the control versus coached group sampling has been appropriately conducted and the extent to which the researchers involved have the training to carry out such research and interpret outcomes.

Considering our earlier doubts about ROI, we suggest alternative ways to measure organizational-level EC outcomes. ROI studies (e.g. Phillips, 2007) combine hard data with estimates and/or self-reporting to produce a monetary value or percentage of improved performance. Instead of ROI, measures of organizational-level impact can be measurements relating to customer service (e.g. customer satisfaction, response time), project management (e.g. deadline and deliverable achievement), productivity and quality (Fairhurst, 2007).

Finally, meta-analysis helps in identifying field-level trends on outcomes and requires review of varying research styles and designs. As more experimental studies are produced we anticipate researchers to produce more meta-analyses on EC outcomes. However, as we experienced here, when a field, such as EC, presents such a variation in the quality of research and the types of EC approaches and empirical questions addressed, a systematic review could bridge different research approaches and identify emerging research themes.

Implications for research and practice based on evidence of coaching outcomes (2b)

The journey rather than the destination: towards a more context-sensitive research agenda

The latter part of our review focused on the contextual factors that affect EC outcomes. We conclude that the field's main gaps are not so much regarding whether coaching works and which types of outcomes are stronger since these are areas that have been tackled in recent review studies (Theeboom et al., 2014; Jones et al., 2016; Grover & Furnham, 2016). Instead, where we still know very little is how the social context influences the process and outcomes of EC and why.

Having summarised what the current research has identified as contextual influences, we use the last column of Table 3, to propose a list of new research questions, per contextual factor, that future EC outcome studies could address. Such questions focus on helping better understand the "journey" of EC, that is, the 'how' and 'why's of the intervention rather that the 'what'. We also propose methods that researchers could employ to investigate each new research question. The aim is to offer to EC scholars examples of research questions and proposed research methods for the design of EC outcome studies that are more attuned to social contextual considerations. For example, we know that several studies have tested whether a specific coaching model or method works (e.g. Mackie's (2014) experimental study on the effectiveness of strength-based coaching in enhancing transformational leadership). What we do not know is how the different coaching models compare to one another in terms of outcomes produced or whether the type of coaching method is less important compared to other elements of the intervention as well as what these are. For instance, De Haan et al. (2011) found that coaching effectiveness is less correlated with technique or intervention than with factors common to all coaching, such as the coaching relationship, positive expectations etc. Studies that will further explore such an important finding or identify variations in coaching models' effectiveness are important. Here, we propose the use of an RCS design to control for coaching model differences.

Another research could explore the impact of coachee's job rank on EC outcomes. Bowles et al. (2007; repeated measures design) found that coaching had strongest performance impact on middle managers and their subordinates than on executives. A study could explore how the EC of senior leaders should differ from that of middle managers to be more effective. A research team could run multiple, multi-level case studies or RCSs to compare a senior executives group with a middle managers group and a control group and unpack differences.

Overall, we see much value in more context-sensitive research and we have indicated the richness of information than can be drawn from such an approach that attends to contextual influences on EC outcomes.

Reflective questions for better research designs. The field's scholarly research on the role of social context in EC outcomes is still at its infancy. This is not at all surprising given that in the wider field of organizational studies, context has been mostly treated in an ad hoc fashion or oriented towards a particular aspect of context (Johns, 2006). The complexity with social context arises from the fact that individuals can both shape and are shaped by the context they belong to (Giddens, 1984), whether that is their organization or the broader environment they operate in. As such, context, within which leadership evolves as a social and goal-oriented influence process, involves spatial and temporal elements (Fischer et al., 2017). It can be an event, a shaper of meaning as well as a bundle of stimuli (Johns, 2006). It is therefore helpful for researchers to think about their treatment of context in study design. The concept of context is important in the complex and dynamic setting of coaching. In Table 3 we have set out some of the facets of the social context that we have found in operation in this setting in the existing literature. We are not arguing that researchers need to be attentive to all of these facets, but to be mindful of their existence and acknowledge the possibilities of influence in their design and analysis plans. We provide here a list of reflections that researchers in this field may need to consider in light of our study findings about context:

a) What are the contextual facets at play that influence the research question? How is the researcher defining context? As an event? An organizational space? A constant? A shaper of meaning? What are the limits being placed around the range or boundaries of context in the research?
b) What questions best help uncover the role of context as a mediator in the relationships under study?
c) What are the potential impacts of context on the action under scrutiny?
d) How might a design capture the more processual elements of coaching, that is, the journey? How is the journey over time accounted for? What is changing in the coach-coachee relationship and their individual contexts that might impact that journey? How can we analyze processual data (see Langley, 1999)?

Obviously, depending on the research epistemology and expertise that each researcher carries, a different set of considerations would need to be accounted for. Quantitative researchers would need to clearly delineate the causal pathways of the contextual variables included in the study for hypothesis testing so as to avoid endogeneity (Antonakis, Bendahan, Jacquart and Lalive, 2010). On the other hand, qualitative research could "expand and sometimes retest empirically-supported theories with qualitative methods to establish causal mechanisms that are not well suited to quantitative testing and to uncover what has changed as well as what has remained the same" (Bluhm, Harman, Lee and Mitchell, 2011, p. 1870). A well-crafted case study research for instance, would involve "careful justification of theory building, theoretical sampling of cases, interviews that limit informant bias, rich presentation of evidence in tables and appendixes, and clear statement of theoretical arguments" (Eisenhardt & Graebner, 2007, p. 30).

Irrespective of method employed, though, as we saw, a study is only as good as its research design and analysis. The EC outcome research can benefit from knowledge drawn from a field that has longer tradition in social contextual analysis.

EC as a process: lessons from the organizational change literature. According to Johns (2006, p. 389) the influence that context has on organizational life is "often unrecognized or underappreciated" and when context is studied, the "contextual features are often studied in a piecemeal fashion, in isolation from each other". Social contextual influences must be studied as they interact and affect an EC intervention.

The stream of research in organization studies that has a long trajectory of looking into the role of context on processes is that of organizational change. EC is, itself, a change process embedded in the organizational context. It is a "complex and demanding process" with multidimensional interrelationships among the individual, the organization and the consultant, guided by four premises: the role of the unconscious in individual and group behavior; the interaction between the individual and the organization; multilevel organizational forces; and the consultant's use of self as tool (Orenstein, 2002, p. 355). Under such a framing, the social context of EC is the meshing of influences between the stakeholders and the environment within which they operate, influencing and being influenced by the EC process with repercussions at multiple levels. Therefore, EC outcomes research could well benefit from methods drawn from the organizational change literature:

Gray, Stensaker and Jansen (2012, p. 124) suggest that the organizational change literature need to contemplate on three key questions: "(a) What role does context play in shaping change processes? (b) Whose change voice is being heard, and at what level of analysis? (c) How does the conceptualization of time influence change processes?". Each question applies to EC outcome research, too, if we are to treat EC as a process embedded and in interrelation with its context. "The nature of the context" in which a change happens "may enhance or restrict aspects of the change itself" (Gray, Stensaker & Jansen, 2012, p. 124). Indeed, the data we present in Table 3 indicate that the contextual influences can enhance or restrict the progress of an EC intervention depending on how these contextual factors are treated by the stakeholders.

Process research methods, embraced by the organizational change literature, can be helpful in EC outcome research. In her seminal paper, Langley (1999) provides several strategies for process data analysis.

These range from a narrative strategy (raw data are constructed into a detailed story) to the quantification strategy (in-depth process data are reduced into a set of quantitative time series that can be statistically analysed) and from the alternate templates strategy (alternative interpretations of the same events are proposed in the analysis) to grounded theory (small units of data (incidents) are systematically compared and gradually constructed into "categories" that describe the phenomenon under study). The final two research strategies that Langley (1999) proposed are the visual mapping strategy (visual graphical representations of data) and temporal bracketing (temporal decomposition allows the constitution of comparative units). These strategies cater for a variety of method preferences and research designs, from quantitative to qualitative. They have not been extensively embraced in EC outcome studies, yet, though.

Two recurring themes in recent organizational change literature, which the EC field should be mindful of, are the importance of time and timing of the intervention and stakeholder voice. In the organizational literature, one finds multiple conceptions of time: clock time (i.e., "the actual passage of quantitative time"), psychological time (i.e., "perceptions of the past and future in the present moment") and socialized time (i.e., "the patterns and temporal ordering associated with social processes and sets of events") (Gray, Stensaker & Jansen, 2012, p. 128). The study of time is important as we need to be "cognizant of how time is being conceptualized by different actors involved in the change (and even by the researchers studying the change), and the ways in which these conceptualizations may influence the change processes and outcomes being studied" (Gray, Stensaker & Jansen, 2012, p. 129).

The second aspect is about the voice of stakeholders. Change research is now increasingly focusing on "whose story they are telling", with alternative voices, from multiple levels of analysis, being invited since decisions at one level trigger responses at different levels and "voice at one level may suppress or enable voice at another level" (Gray, Stensaker and Jansen, 2012, pp. 126–127). A lesson to be drawn for the EC outcome research is the design of studies that are more stakeholder inclusive. As presented in Table 1, only 4 EC outcome studies focused on all three EC stakeholders - an area that future research should address.

Evidence-based recommendations for practitioners

Practitioners can benefit from our research by studying the range of EC outcomes identified and by exploring the full potential for positive change that an EC intervention is observed to have (Table 2). Practitioners could also benefit by the identified EC pitfalls of Table 2 as well as the contextual themes summarised in Table 3 and explore how they can positively manage contextual influences to improve EC practice or mitigate any negative influences.

This study sought to shift the field's attention to the role of the social context and the power it has to transform EC practice. Other reflections for coaching practitioners to consider are:

a. Prior to a coaching session, coaches should develop a curiosity about the social context of which the coachee is a part. Coaches could inquire into the organizational culture and in particular what rituals and routines, control systems, structures and power relationships are serving or impeding organizational action. These are likely to have an effect on the EC intervention outcomes and their sustainability. Also, research suggests cultures can be nudged (Ravasi & Schultz, 2006), therefore, the coach should be mindful of the potential that each coachee has to change cultures to help bring about progress. Coach's in-depth understanding of the organizational context can bring outcomes that cut across contextual levels and connect the micro (individual) with the *meso* (group/departmental) and *macro* (organizational) level.
b. Stewart (1982) suggested that leadership jobs are flexible spaces, however they are comprised of demands, that is to say, the task that must be done to survive in the organization. Are these contextual constraints for the coachee or are they choices? A coach could explore the demands of the job and whether they are real or imagined. Reframing these demands as choices would then require to work with the coachee and determine what choices are being made and whether these are serving the coachee and the organization.
c. The framing of the EC session is critical in shaping the experience of the session (Table 3). Aside from the usual EC contracting principles, it is helpful to explore more fully coaches' views on leadership as these are likely to influence the type of EC they provide. Do these align with the coachee's and the sponsoring organization's views on leadership? By surfacing these assumptions, it becomes easier for the coach to have more meaningful discussions of the coachee's and sponsoring organization's practices and contextual challenges.
d. In EC the use of psychometrics can be helpful in casting a light on leadership practices and style. Several EC studies use data from 360-degree feedback. The use of such feedback must be carefully and developmentally applied. Sourcing 360-degree feedback about what others appreciate about the coachee -and what they would like to see more of- may have a more empowering effect on the coachee for positive change than a focus on weaknesses alone.
e. EC is a leadership development intervention and as such entails the mastery of new leadership skills or existing ones. Understanding the coachee's motivation and approach to seeking knowledge has been shown to influence leadership practice (Fischer et al., 2015). Specifically, research on achievement goals has found that *mastery* goals positively impact achievement-related outcomes, but in academic contexts the social desirability (i.e. goals perceived as nice) and social utility (i.e. high probability of success) of mastery goals play a moderating role on achievement outcomes (Dompnier, Darnon & Butera, 2009). Coaches could usefully explore how coachees seek knowledge and how their perceived social desirability and social utility of coaching goals may affect the EC outcomes. This calls for self-reflection by the coaches on what their own approach to skills mastery and reflection is.
f. Only four studies have looked into comparing the outcomes of EC with other leadership development interventions or no intervention. EC has been compared with a traditional management training program (Olivero, Bane & Kopelman (1997), peer- and self-coaching (Sue-Chan & Latham, 2004) and coaching-based leadership programmes (i.e. internal coaching) (Moen & Skaalvik, 2009; Jones, Woods & Guillaume, 2016). With the exception of Jones, Woods and Guillaume's (2016) study who found that coaching was more effective when conducted by internal coaches, all other studies found EC to be more effective than any other intervention it has been compared with. While we need to understand better why EC tends to be more effective on average than these other interventions, HR professionals, executive coaches and management educators may collaborate to explore ways by which coaching can be better integrated within other leadership development initiatives. This may produce win-win outcomes which are stronger than the outcomes produced from EC alone.

Conclusions

Day (2000, p. 582) observed a "relative dearth" of scholarly studies on leadership development and a "disconnection" between the practice of leadership development and its scientific foundation. We have attempted here to make some connections between the two within the context of EC research and practice. We made several contributions:

First, we assessed the quality of EC research and reviewed the evidence on EC outcomes. Secondly, we built on this assessment and went against the grain of the EC scholarly work to criticize the field's almost obsessive focus on the "end" or "destination" (i.e. what the EC outcomes are and how strong they are) at the expense of the "journey" (what EC involves as a practice and in what ways the social context within which it takes place matters to this journey). EC is not going

away any time soon even if the evidence that it works is not yet (or will ever be) strong enough. It is part of human nature to invest in "hope" rather than solid "outcomes". Coachees and organizations are no different. By drawing the field's attention more on the nature and social contextual influences of EC interventions this helps to develop more context-sensitive and informed interventions and creates ultimately a shortcut to better outcome research. Thirdly -and related to our second contribution- we suggest reframing EC as a social rather than an individual intervention - one where the organization, the coach and the coachee co-create new meanings embedded and shaped by the social context within which the intervention is applied. The final component of our contribution is our suggestion of a future research and practice agenda in response to our proposed reframing of the field. We argue that the field could benefit from knowledge drawn from the organizational change literature and process research strategies.

Our findings aim at mixed audiences of *practitioners* (e.g. professional coaches, HR, leadership development and organizational development professionals as well as coachees at different career stages) and *scholars* (EC, leadership, leadership development, management/executive education, organizational behavior, psychology and professional training). EC research can benefit from such inter- and multi-disciplinarity. Our paper has broader implications for leadership development theory and practice in light of recent calls for more evidence-based leadership development practices (Bartunek, 2012; Klimoski & Amos, 2012) as well as recent trends in management education for more customized training that facilitates reflection and personal development (see Datar et al., 2010; Petriglieri et al., 2011) and is more context-relevant (Armstrong & Sadler-Smith, 2008, p. 571). To that end, we echo Grover and Furnham's (2016, p. 36) suggestion that the way forward for better EC outcome research is the development of independent working groups of coaches, academics, sponsoring organizations and any other stakeholders that have an interest in EC research to offer best practices guidelines and recommendations for more rigorous research.

Conflicts of interest declaration

No conflict of interest.

Funding declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment

We would like to thank Andrew White (Associate Dean for Executive Education at the Saïd Business School) for providing us with support to carry out this research. We are also very grateful to the associate editor, David Day, and the anonymous reviewers for the very constructive feedback they provided throughout the revisions of this manuscript.

References[1]

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*(2), 334–345.
Anderson, J. P. (2002). Executive coaching and REBT: Some comments from the field. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 20*(3–4), 223–233.
Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.
Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086–1120.
Arbaugh, J. B. (2011). From the editors: It takes a village to build a field. *Academy of Management Learning & Education, 10*(4), 557–560.
Armstrong, S. J. (2011). From the editors: Continuing our quest for meaningful impact on management practice. *Academy of Management Learning & Education, 10*(2), 181–187.
Armstrong, S. J., & Sadler-Smith, E. (2008). Learning on demand, at your own pace, in rapid bite-sized chunks: The future shape of management development? *Academy of Management Learning & Education, 7*(4), 571–586.
Athanasopoulou, A., & Dopson, S. (2015). *Developing leaders by executive coaching: Practice and evidence.* Oxford: Oxford University Press.
Baron, L., & Morin, L. (2009). The coach-coachee relationship in executive coaching: A field study. *Human Resource Development Quarterly, 20*(1), 85–106.
Bartunek, J. (2012). Introduction: Do we as academic instructors use academic evidence in developing leaders? *Academy of Management Learning & Education, 11*(4), 684.
Ben-Hador, B. (2016). Coaching executives as tacit performance evaluation: A multiple case study. *Journal of Management Development, 35*(1), 75–88.
Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. *Journal of Management Studies, 48*, 1866–1891.
Bono, J. E., Purvanova, R. K., Towler, A. J., & Peterson, D. B. (2009). A survey of executive coaching practices. *Personnel Psychology, 62*(2), 361–404.
Bowles, S. V., & Picano, J. J. (2006). Dimensions of coaching related to productivity and quality of life. *Consulting Psychology Journal: Practice & Research, 58*(4), 232–239.
Bowles, S., Cunningham, C. J. L., De La Rosa, G. M., & Picano, J. (2007). Coaching leaders in middle and executive management: Goals, performance, buy-in. *Leadership and Organization Development Journal, 28*(5), 388–408.
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*(3), 277–294.
Bozer, G., Sarros, J. C., & Santora, J. C. (2014). Academic background and credibility in executive coaching effectiveness. *Personnel Review, 43*(6), 881–897.
Bozer, G., Baek-Kyoo, J., & Santora, J. C. (2015). Executive coaching: Does coach-coachee matching based on similarity really matter? *Consulting Psychology Journal: Practice & Research, 67*(3), 218–233.
Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71*(2), 179–184.
Coutu, D., Kauffman, C., Charan, R., Peterson, D. B., Maccoby, M., Scoular, P. A., & Grant, A. M. (2009). HBR research report: What can coaches do for you? *Harvard Business Review, 87*(1), 91–97.
Cowan, K. (2013). What are the experiences of external executive coaches working with coachees' assigned goals? *International Journal of Evidence Based Coaching and Mentoring, 7*, 14–25.
Datar, S. M., Garvin, D. A., & Cullen, P. G. (2010). *Rethinking the MBA: Business education at a crossroads.* Boston, MA: Harvard Business Press.
Dawdy, G. N. (2004). *Executive coaching: A comparative design exploring the perceived effectiveness of coaching and methods. Unpublished doctoral dissertation.* Minneapolis, MN: Capella University.
Day, D. V. (2000). Leadership development: A review in context. *The Leadership Quarterly, 11*(4), 581–613.
Day, D. V. (2014). Time and leadership. *How time impacts groups, organizations and methodological choices: . Vol. 2. Time and work* (pp. 30–). London, UK: Psychology Press.
Day, D. V., & Lord, R. G. (1988). Executive leadership and organizational performance: Suggestions for a new theory and methodology. *Journal of Management, 14*, 453–464.
De Haan, E., Bertie, C., Day, A., & Sills, C. (2010). Clients' critical moments of coaching: Toward a "client model" of executive coaching. *Academy of Management Learning & Education, 9*(4), 607–621.
De Haan, E., Culpin, V., & Curd, J. (2011). Executive coaching in practice: What determines helpfulness for clients of coaching? *Personnel Review, 40*(1), 24–44.
De Haan, E., & Duckworth, A. (2013). Signalling a new trend in executive coaching outcome research. *International Coaching Psychology Review, 8*(1), 6–19.
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 & Research, 65*(1), 40–57.
De Haan, E., & Nieß, C. (2012). Critical moments in a coaching case study: Illustration of a process research model. *Consulting Psychology Journal: Practice & Research, 64*(3), 198–224.
Denis, J.-L., Langley, A., & Sergi, S. (2012). Leadership in the plural. *The Academy of Management Annals, 6*(1), 211–283.
Denzin, N. (1984). *The research act.* Englewood Cliffs, NJ: Prentice Hall.
Diedrich, R. C. (1996). An interactive approach to executive coaching. *Consulting Psychology Journal: Practice & Research, 48*(2), 61–66.
Diochon, P. F., & Nizet, J. (2015). Ethical codes and executive coaches: One size does not fit all. *The Journal of Applied Behavioral Science, 51*(2), 277–301.
Dompnier, B., Darnon, C., & Butera, F. (2009). Faking the desire to learn: A clarification of the link between mastery goals and academic achievement. *Psychological Science, 20*(8), 939–943.
Drake, D. (2008). Finding our way home: Coaching's search for identity in a new era. *Coaching: An International Journal of Theory, Research and Practice, 1*(1), 15–26.
Drake, D. B. (2009). Evidence is a verb: A relational approach to knowledge and mastery in coaching. *International Journal of Evidence Based Coaching and Mentoring, 7*(1), 1–12.
Edwards, A. L. (1990). Construct validity and social desirability. *American Psychologist, 45*, 287–289.
Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550.
Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*, 25–32.

[1] Note: The list of the 110 papers that were included in the systematic review are available upon request.

Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*(2), 155–166.
Ely, K., Boyce, L. A., Zaccaro, S. J., Hernez-Broome, G., & Whyman, W. (2010). Evaluating leadership coaching: A review and integrated framework. *The Leadership Quarterly, 21*, 585–599.
Ennis, S. A., Goodman, R., Otto, J., & Stern, L. R.. The executive coaching handbook: Principles and guidelines for a successful coaching partnership (4th ed.). (2008). http://www.executivecoachingforum.com, Accessed date: May 2011 ((with Hodgetts, W., Hunt, J. M., & R. Mansfield) (Vitti, M., & Yahanda, N. participated in the 1st edition)).
Fairhurst, P. (2007). Measuring the success of coaching. *Training Journal, 7*, 52–55.
Feldman, D. C., & Lankau, M. J. (2005). Executive coaching: A review and agenda for future research. *Journal of Management, 31*(6), 829–848.
Fischer, M., Dopson, S., Ferlie, E., & McGivern, G. (2015). Knowledge leadership: Mobilising management research by becoming a knowledge object. *Human Relations, 69*(7), 1563–1585.
Fischer, R. L., & Beimers, D. (2009). "Put me in, coach": A pilot evaluation of executive coaching in the nonprofit sector. *Nonprofit Management & Leadership, 19*(4), 507–522.
Fischer, T., Dietz, J., & Antonakis, J. (2017). Leadership process models: A review and synthesis. *Journal of Management, 43*(6), 1726–1753.
Foster, S., & Lendl, J. (1996). Eye movement desensitization and reprocessing: Four case studies of a new tool for executive coaching and restoring employee performance after setbacks. *Consulting Psychology Journal: Practice and Research, 48*(3), 155–161.
Freedman, A. M., & Perry, J. A. (2010). Executive consulting under pressure: A case study. *Consulting Psychology Journal: Practice and Research, 62*(3), 189–202.
Gan, G. C., & Chong, C. W. (2015). Coaching relationship in executive coaching: A Malaysian study. *Journal of Management Development, 34*(4), 476–493.
Garman, A. N., Whiston, D. L., & Zlatoper, K. W. (2000). Media perceptions of executive coaching and the formal preparation of coaches. *Consulting Psychology Journal: Practice and Research, 52*(3), 201–205.
Gaskell, C., Logan, J., & Nicholls, L. (2012). Igniting business performance through transformational coaching. *Strategic HR Review, 11*(1), 18–24.
Gegner, C. (1997). *Coaching: Theory and practice. Unpublished master's dissertation.* San Francisco, CA: University of San Francisco.
Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Cambridge: Polity Press.
Gorringe, A. (2011). The place for coaching for clinical leadership. *International Journal of Clinical Leadership, 17*(1), 19–23.
Grant, A. M. (2003). *Evidence-based coaching: The what, how and why. Paper presented at the first Australian evidenced-based coaching conference.* Australia: Sydney University.
Grant, A. M. (2011). *Workplace, executive and life coaching: An annotated bibliography from the behavioural science and business literature (1st Jan 2011).* Australia: Coaching Psychology Unit, University of Sydney.
Grant, A. M. (2012a). An integrated model of goal-focused coaching: An evidence-based framework for teaching and practice. *International Coaching Psychology Review, 7*(2), 146–165.
Grant, A. M. (2012b). 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. (2014). The efficacy of executive coaching in times of organisational change. *Journal of Change Management, 14*(2), 258–280.
Grant, A. M., Curtayne, L., & Burton, G. (2009). Executive coaching enhances goal attainment, resilience and workplace wellbeing: A randomised controlled study. *The Journal of Positive Psychology, 4*(5), 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: Practice and Research, 62*(3), 151–168.
Grant, A. M., & Zackon, R. (2004). Executive, workplace and life coaching: Findings from a large-scale survey of international coach federation members. *International Journal of Evidence Based Coaching and Mentoring, 2*(2), 1–15.
Gray, B., Stensaker, I. G., & Jansen, K. J. (2012). Qualitative challenges for complexifying organizational change research: Context, voice, and time. *The Journal of Applied Behavioral Science, 48*(2), 121–134.
Gray, D., Burls, A., & Kogan, M. (2014). Salutogenisis and coaching: Testing a proof of concept to develop a model for practitioners. *International Journal of Evidence Based Coaching & Mentoring, 12*(2), 41–58.
Gray, D. E., Ekinci, Y., & Goregaokar, H. (2011a). Coaching SME managers: Business development or personal therapy? A mixed methods study. *International Journal of Human Resource Management, 22*(4), 863–882.
Gray, D. E., Gabriel, Y., & Goregaokar, H. (2015). Coaching unemployed managers and professionals through the trauma of unemployment: Derailed or undaunted? *Management Learning, 46*(3), 299–316.
Greif, G. (2007). Advances in research on coaching outcomes. *International Coaching Psychology Review, 2*(3), 222–247.
Griffiths, K. E., & Campbell, M. A. (2008). Regulating the regulators: Paving the way for international, evidence-based coaching standards. *International Journal of Evidence Based Coaching and Mentoring, 6*(1), 19–31.
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), 1–41.
Hall, D. T., Otazo, K. L., & Hollenbeck, G. P. (1999). Behind closed doors: What really happens in executive coaching. *Organizational Dynamics, 27*(3), 39–53.
Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist, 43*, 621–626.
Hooijberg, R., & Lane, N. (2009). Using multisource feedback coaching effectively in executive education. *Academy of Management Learning & Education, 8*(4), 483–493.
Hoon, C. (2013). Meta-synthesis of qualitative case studies: An approach to theory building. *Organizational Research Methods, 16*(4), 522–556.
Howard, A. R. (2015). Coaching to vision versus coaching to improvement needs: A preliminary investigation on the differential impacts of fostering positive and negative emotion during real time executive coaching sessions. *Frontiers in Psychology, 6*, 455.
ICF (2012). The 2012 ICF global coaching study. Retrieved from www.coachfederation.org/coachingstudy2012.
ICF (2014). The 2014 ICF Global consumer awareness study. Retrieved from http://coachfederation.org/consumerstudy2014.
Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review, 31*(2), 386–408.
Jones, R. A., Rafferty, A. E., & Griffin, M. A. (2006). The executive coaching trend: Towards more flexible executives. *Leadership and Organization Development Journal, 27*(7), 584–596.
Jones, R. J., Woods, S. A., & Guillaume, Y. 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.
Judge, W. Q., & Cowell, J. (1997). The brave new world of executive coaching. *Business Horizons, 40*(4), 71–77.
Kampa-Kokesch, S., & Anderson, M. Z. (2001). Executive coaching: A comprehensive review of the literature. *Consulting Psychology Journal: Practice and Research, 53*, 205–228.
Kauffman, C., & Hodgetts, W. H. (2016). Model agility: Coaching effectiveness and four perspectives of a case study. *Consulting Psychology Journal: Practice & Research, 68*(2), 157–176.
Kiel, F., Rimmer, E., Williams, K., & Doyle, M. (1996). Coaching at the top. *Consulting Psychology Journal: Practice & Research, 48*(2), 67–77.
Kilburg, R. R. (1996). Toward a conceptual understanding and definition of executive coaching. *Consulting Psychology Journal: Practice and Research, 48*(2), 134–144.
Klag, M., & Langley, A. (2013). Approaching the conceptual leap in qualitative research. *International Journal of Management Reviews, 15*(2), 149–166.
Klimoski, R., & Amos, B. (2012). Practicing evidence-based education in leadership development. *Academy of Management Learning & Education, 11*(4), 685–702.
Kombarakaran, F. A., Yang, J. A., Baker, M. N., & Fernandes, P. B. (2008). Executive coaching: It works!. *Consulting Psychology Journal: Practice and Research, 60*(1), 78–90.
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134.
Ladegard, G., & Gjerde, S. (2014). Leadership coaching, leader role-efficacy, and trust in subordinates. A mixed methods study assessing leadership coaching as a leadership development tool. *The Leadership Quarterly, 25*(4), 631–646.
Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review, 24*(4), 691–710.
Laske, O. E. (1999). *Transformative effects of coaching on executive's professional agenda.* Ann Arbor, MI: University of Michigan Press.
Laske, O. E. (2004). Can evidence-based coaching increase ROI? *International Journal of Evidence-Based Coaching and Mentoring, 2*(2), 41–53.
Lawrence, P. (2015). Building a coaching culture in a small Australian multinational organisation. *Coaching: An International Journal of Theory, Research and Practice, 8*(1), 53–60.
Lee, T. W. (1999). *Using qualitative methods in organizational research.* Thousand Oaks, CA: Sage.
Leedham, M. (2005). The coaching scorecard: A holistic approach to evaluating the benefits of business coaching. *International Journal of Evidence Based Coaching and Mentoring, 3*(2), 30–44.
Levenson, A. (2009). Measuring and maximizing the business impact of executive coaching. *Consulting Psychology Journal: Practice and Research, 61*(2), 103–121.
Lewis-Duarte, M., & Bligh, M. C. (2012). Agents of "influence": Exploring the usage, timing, and outcomes of executive coaching tactics. *Leadership and Organization Development Journal, 33*(3), 255–281.
Luthans, F., & Peterson, S. (2003). 360-Degree feedback with systematic coaching: Empirical analysis suggests a winning combination. *Human Resource Management, 42*(3), 243–256.
MacKie, D. (2014). The effectiveness of strength-based executive coaching in enhancing full range leadership development: A controlled study. *Consulting Psychology Journal: Practice & Research, 66*(2), 118–137.
Martineau, J. W., & Patterson, T. E. (2010). In E. Van (Ed.). *Evaluating leader development.*
Michelman, P. (2004). Do you need an executive coach? *Harvard Management Update, 9*(12), 3–4.
Moen, F., & Allgood, E. (2009). Coaching and the effect on self-efficacy. *Organization Development Journal, 27*(4), 69–81.
Moen, F., & Federici, R. A. (2012a). The effect from external executive coaching. *Coaching: An International Journal of Theory, Research and Practice, 5*(2), 113–131.
Moen, F., & Federici, R. A. (2012b). The effect of external executive coaching and coaching-based leadership on need satisfaction. *Organization Development Journal, 30*(3), 63–74.
Moen, F., & Skaalvik, E. (2009). The effect from executive coaching on performance psychology. *International Journal of Evidence Based Coaching and Mentoring, 7*(2), 31–49.
Nicholson, R., & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist, 45*(2), 290–292.
Nikolova, N., Clegg, S., Fox, S., Bjørkeng, K., & Pitsis, T. (2013). Uncertainty reduction through everyday performative language work. *International Studies of Management & Organization, 43*(3), 74–89.

Olivero, G., Bane, K. D., & Kopelman, R. E. (1997). Executive coaching as a transfer of training tool: Effects on productivity in a public agency. *Public Personnel Management, 26*(4), 461–469.
O'Neill, M. B. (2007). *Executive coaching with backbone and heart: A systems approach to engaging leaders with their challenges* (2nd ed.). San Francisco, CA: Jossey-Bass.
Orenstein, R. L. (2002). Executive coaching: It's not just about the executive. *The Journal of Applied Behavioral Science, 38*(3), 355–374.
Osatuke, K., Yanovsky, B., & Ramsel, D. (2016). Executive coaching: New framework for evaluation. *Consulting Psychology Journal: Practice and Research.* http://dx.doi.org/10.1037/cpb0000073.
Packard, T., & Jones, L. (2015). An outcomes evaluation of a leadership development initiative. *Journal of Management Development, 34*(2), 153–168.
Parker-Wilkins, V. (2006). Business impact of executive coaching: Demonstrating monetary value. *Industrial and Commercial Training, 38*(3), 122–127.
Passmore, J. (2009). Seeing beyond the obvious: Executive coaching and I-O psychologists. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*(3), 272–276.
Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review - A new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy, 10*(1), 21–34.
Peterson, D. B. (1993a). *Skill learning and behavior change in an individually tailored management coaching and training program. Doctoral dissertation.* University of Minnesota.
Peterson, D. B. (1993b). *Measuring change: A psychometric approach to evaluating individual training outcomes. Symposium conducted at the Eighth Annual Conference of the Society for Industrial and Organizational Psychologists, San Francisco, CA*.
Peterson, D. B., & Millier, J. (2005). The alchemy of coaching: “You're good, Jennifer, but you could be really good”. *Consulting Psychology Journal: Practice & Research, 57*(1), 14–40.
Petriglieri, G., Wood, J., & Petriglieri, J. (2011). Up close and personal: Building foundations for leaders' development through the personalization of management learning. *Academy of Management Learning & Education, 10*(3), 430–450.
Pettigrew, A. M. (1985). *The awakening giant: Continuity and change in ICI.* Oxford: Basil Blackwell.
Phillips, J. J. (2007). Measuring the ROI of a coaching intervention, Part 2. *Performance Improvement, 46*(10), 10–23.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*(1), 539–569.
Polsfuss, C., & Ardichvili, A. (2008). Three principles psychology: Applications in leadership development and coaching. *Advances in Developing Human Resources, 10*(5), 671–685.
Pratt, M. G. (2008). Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods, 11*(3), 481–509.
Ravasi, D., & Schultz, M. (2006). Responding to organizational identity threats: Exploring the role of organizational culture. *Academy of Management Journal, 49*(3), 433–458.
Rynes, S. L., & Brown, K. G. (2011). Where are we in the “long march to legitimacy?” Assessing scholarship in management learning and education. *Academy of Management Learning & Education, 10*(4), 561–582.
Salter, T. (2014). Mentor and coach: Disciplinary, interdisciplinary and multidisciplinary approaches. *International Journal of Evidence Based Coaching & Mentoring, 8*, 1–8.
Sammut, K. (2014). Transformative learning theory and coaching: Application in practice. *International Journal of Evidence Based Coaching & Mentoring, 8*, 39–53.
Segers, J., Vloeberghs, D., Henderickx, E., & Inceoglu, I. (2011). Structuring and understanding the coaching industry: The coaching cube. *Academy of Management Learning & Education, 10*(2), 204–221.
Shamir, B. (2011). Leadership takes time: Some implications of (not) taking time seriously in leadership research. *The Leadership Quarterly, 22*, 307–315.
Sherpa Consulting (2014). Executive coaching survey '14: Evidence & interaction. http://www.sherpacoaching.com/pdf%20files/2014%20Executive%20Coaching%20Survey%20-%20Public%20report.pdf.
Siemsen, E., Roth, A., & Oliveira, P. (2010). Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods, 13*(3), 456–476.
Simonton, D. K. (2003). Qualitative and quantitative analyses of historical data. *Annual Review of Psychology, 54*, 617–640.
Smith, I. M., & Brummel, B. J. (2013). Investigating the role of the active ingredients in executive coaching. *Coaching: An International Journal of Theory, Research and Practice, 6*(1), 57–71.
Smither, J. W., London, M., Flautt, R., Vargas, Y., & Kucine, I. (2003). Can working with an executive coach improve multisource feedback ratings over time? A quasi-experimental field study. *Personnel Psychology, 56*(1), 23–44.
Sonesh, S. C., Coultas, C. W., Marlow, S. L., Lacerenza, C. N., Reyes, D., & Salas, E. (2015). Coaching in the wild: Identifying factors that lead to success. *Consulting Psychology Journal: Practice & Research, 67*(3), 189–217.
Stober, D. R., & Grant, A. M. (2006). *Evidence based coaching handbook: Putting best practices to work for your clients.* Hoboken, NJ: John Wiley & Sons.
Stake, R. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.
Stevens, J. H., Jr. (2005). Executive coaching from the executive's perspective. *Consulting Psychology Journal: Practice and Research, 57*(4), 274–285.
Stewart, R. (1982). *Choices for the manager.* London: Prentice Hall.
Sturm, R. E., & Antonakis, J. (2015). Interpersonal power: A review, critique, and research agenda. *Journal of Management, 41*(1), 136–163.
Styhre, A. (2008). Coaching as second-order observations: Learning from site managers in the construction industry. *Leadership and Organization Development Journal, 29*(3), 275–290.
Sue-Chan, C., & Latham, G. P. (2004). The relative effectiveness of external, peer, and self-coaches. *Applied Psychology. An International Review, 53*(2), 260–278.
Tamir, L. M., & Finfer, L. A. (2016). Executive coaching: The age factor. *Consulting Psychology Journal: Practice & Research, 68*(4), 313–325.
Thach, E. C. (2002). The impact of executive coaching and 360 feedback on leadership effectiveness. *Leadership and Organization Development Journal, 23*(4), 205–214.
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. *The Journal of Positive Psychology, 9*(1), 1–18.
Toogood, K. (2012). Strengthening coaching: An exploration of the mindset of executive coaches using strengths-based coaching. *International Journal of Evidence Based Coaching and Mentoring, 6*, 72–87.
Turner, C., & McCarthy, G. (2015). Coachable moments: Identifying factors that influence managers to take advantage of coachable moments in day-to-day management. *International Journal of Evidence Based Coaching & Mentoring, 13*(1), 1–13.
Van Diemen van Thor, F. (2014). Using the ‘autodidact’ subject-object interview in coaching: The experience of learning to administer and score the subject-object interview through self-teaching. *International Journal of Evidence Based Coaching & Mentoring, 8*, 9–23.
Wales, S. (2003). Why coaching? *Journal of Change Management, 3*(3), 275–282.
Ward, G., Van de Loo, E., & ten Have, S. (2014). Psychodynamic group executive coaching: A literature review. *International Journal of Evidence Based Coaching & Mentoring, 12*(1), 63–78.
Wasylyshyn, K. M. (2003). Executive coaching: An outcome study. *Consulting Psychology Journal: Practice & Research, 55*(2), 94–106.
Wasylyshyn, K. M., Gronsky, B., & Haas, J. W. (2006). Tigers, stripes, and behavior change: Survey results of a commissioned coaching program. *Consulting Psychology Journal: Practice and Research, 58*(2), 65–81.
Winum, P. C. (2005). Effectiveness of a high-potential African American executive: The anatomy of a coaching engagement. *Consulting Psychology Journal: Practice and Research, 57*(1), 71–89.
Wise, P. S., & Voss, L. S. (2002). *The case for executive coaching, Research Report from Lore Research Institute.* 1–13.
Witherspoon, R., & White, R. P. (1996). Executive coaching: A continuum of roles. *Consulting Psychology Journal: Practice and Research, 48*(2), 124–133.
Yin, R. (1994). *Case study research: Design and methods.* Thousand Oaks, CA: Sage.
Zarecky, A. (2014). How strengths-focussed coaching can help military personnel in their transition to ‘civvy street’. *International Journal of Evidence Based Coaching & Mentoring, 8*, 54–66.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75–98.