The Leadership Quarterly 21 (2010) 585–599

Contents lists available at ScienceDirect

The Leadership Quarterly

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

Evaluating leadership coaching: A review and integrated framework

Katherine Ely[a,*], Lisa A. Boyce[b], Johnathan K. Nelson[a], Stephen J. Zaccaro[a], Gina Hernez-Broome[c], Wynne Whyman[d]

[a] George Mason University
[b] United States Air Force Academy
[c] University of the Rockies
[d] Leadership Consultant

ARTICLE INFO

Keywords:
Coaching
Training evaluation
Leadership development

ABSTRACT

Leadership coaching reflects an evolving dynamic between the client and coach that is qualitatively different from most approaches to leadership development and therefore holds particular challenges for evaluation. Based on reviews of academic and practitioner literatures, this paper presents an integrated framework of coaching evaluation that includes formative evaluations of the client, coach, client–coach relationship, and coaching process, as well as summative evaluations based on coaching outcomes. The paper also includes a quantitative synthesis examining evaluation methodologies in 49 leadership coaching studies. The results revealed that self-reported changes in clients' leadership behaviors are the most frequently assessed coaching outcome, followed by clients' perceptions of the effectiveness of coaching. Recommendations to advance coaching evaluation research include the creation of collaborative partnerships between the evaluation stakeholders (client, coach, client's organization, and coaching organization) to facilitate systematic formative evaluations, the collection of multi-source and multi-level data, and the inclusion of distal outcomes in evaluation plans.

1. Introduction

Leadership coaching has only recently become a common mode of leadership development (Kampa-Kokesch & Anderson, 2001) with the last decade witnessing unprecedented growth in coaches, coaching programs, and coaching publications (Bolch, 2001; Brotman, Liberi & Wasylyshyn, 1998; Douglas & Morley, 2000; Goldsmith & Lyons, 2005; Quick & Macik-Frey, 2004). While leadership coaching programs have grown to meet the increased demand, little guidance exists on how to evaluate this unique leadership development practice. With this increased prevalence of leadership coaching in organizations – and the substantial time and costs associated with formal leadership coaching – evaluation becomes imperative for both assessing the outcomes of the leadership development intervention (i.e., summative evaluation) and also for improving program development and implementation (i.e., formative evaluation) in order to empirically advance coaching practices.

Leadership coaching is broadly defined in terms of a relationship between a client and a coach that facilitates the client becoming a more effective leader (Douglas & Morley, 2000; Kilburg, 1996; Peterson & Hicks, 1999; Witherspoon & White, 1997). For our research, we capitalize on the Center for Creative Leadership's (CCL) definition of coaching as a formal one-on-one relationship, "in which the coachee and coach collaborate to assess and understand the coachee and his or her leadership developmental tasks, to challenge current constraints while exploring new possibilities, and to ensure accountability and support for reaching goals and sustaining development" (Ting & Hart, 2004, p. 116). In addition to describing the coaching relationship, process, and results, this definition highlights the unique aspects of leadership coaching compared to other leadership development programs or interventions; namely, a focus upon a dynamic one-on-one relationship between a coach and client.

* Corresponding author. George Mason University, Department of Psychology, 4400 University Drive, MSN 3F5, Fairfax, VA 22030, USA.
E-mail address: kely@gmu.edu (K. Ely).

1048-9843/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.leaqua.2010.06.003

Leadership coaching is different from other approaches to leadership development and holds particular challenges for evaluation. Because coaching is provided in the context of a one-on-one relationship and the coaching process varies greatly from person to person, many of the methods of evaluating traditional training interventions may not be as appropriate. Coaching is designed and implemented on an individual level. To address the difficulty inherent in coaching evaluation, our goal is to present a conceptual model of leadership coaching and to present a systematic framework to evaluate it. Researchers and practitioners generally differentiate between two types of evaluation, summative and formative (Brown & Gerhardt, 2002; Patton, 1994). Summative evaluations assess the effectiveness of completed interventions whereas formative evaluations are process oriented and focus on identifying areas for program improvement (Beyer, 1995; Bhola, 1990; Patton, 1994). We propose that evaluating coaching requires an integrated evaluation framework including summative and formative criteria. Additionally, we review the leadership coaching literature—with an emphasis on how coaching is evaluated, to highlight the current state of evaluation practices. Finally, we make recommendations for researchers and practitioners toward promoting effective practices in evaluating leadership coaching.

2. The unique nature of leadership coaching

Leadership coaching differs from traditional leadership development in four ways: (a) leadership coaching focuses on the needs of the individual client as well as the client's organization and the unique characteristics each brings, (b) leadership coaching requires coaches to have unique skill sets, (c) leadership coaching places a premium on the client–coach relationship, and (d) leadership coaching demands process flexibility to achieve desired results. These four (client, coach, client–coach relationship, and coaching process) provide a foundation for understanding the unique nature of leadership coaching, the resulting framework, and implications for its evaluation.

2.1. Client needs

In this article we use the term “client” to refer to individuals formally involved in a coaching relationship for the specific purpose of becoming a more effective leader; coachee is another term used in the literature for this individual. Client needs reflect a variety of inputs to the coaching process, ranging from the reasons for engaging in coaching to the characteristics of clients and their organizations. For example, clients seek the services of leadership coaches for a variety of reasons, such as a desire to increase self-awareness, to develop leadership skills, or to gain insights regarding a current or future situation (Blackman, 2006; Schlosser, Steinbrenner, Kumata, & Hunt, 2006). Underlying these objectives is a need to focus on and improve leadership competencies, behaviors, attitudes, perspectives, and/or strategies that translate into effective job performance and organizational outcomes. Although certain competencies may be more appropriate for development through coaching (e.g., interpersonal skills, communication, self-management, and general leadership skills; Judge & Cowell, 1997; Society for Industrial and Organizational Psychology, 2008), the variety of competencies that can be developed within and across a coaching program may be quite numerous (see Borman & Brush, 1993; Fleishman et al., 1991 for a discussion of leadership competencies).

Further, although clients seek a professional coach for a myriad of reasons, organizations are generally interested in coaching initiatives to meet one or more of the following needs: improve leadership skills to ensure that standards are met (i.e., performance management), develop new leadership skills to meet new or changing job requirements (i.e., change management), develop leadership capabilities for future roles (i.e., talent management), develop leadership capabilities to enhance personnel satisfaction and long-term opportunities (i.e., career management), and ensure leadership breadth and depth for building and leading the organization (i.e., organization management) (Hall, Otazo, & Hollenbeck, 1999; Peterson & Hicks, 1999; Witherspoon & White, 1997). Although the individual and organizational needs are not necessarily incompatible, leadership coaching programs are in the unique position of needing to be responsive to both the individual client and the client's organization.

In addition, each client brings characteristics (e.g., personality, values, beliefs, experiences), including readiness to be coached and commitment to the coaching program. Organizations also bring varying levels of readiness and commitment to coaching initiatives with some organizations providing extensive support and participation, such as providing resources, assessment data, and performance opportunities. Even if a coaching program establishes minimum requirements for readiness (e.g., cognitive readiness such as reading or computer skills; developmental readiness, such as self-efficacy or conscientiousness; and change readiness, such as career exploration or feedback seeking; Boyce, Zaccaro, & Wisecarver, 2010) and commitment (e.g., admitted need, desire to improve performance, willingness to make time and perform work, openness to the experience), the range of attitudes and motivation levels varies between and within clients and their organizations during the developmental experience.

The nature of leadership coaching in meeting the diverse needs of the client is also reflected in the observation that clients engaged in coaching programs may be in different career stages while performing at any level of the organization. With five to seven careers, not just job changes expected in a lifetime (Flower, 2002), leaders' career stages no longer necessarily mirror their organizational responsibilities. Clients participating in leadership coaching bridge organization levels from direct shop floor leaders to strategic corporate executives and as a result present to the coaching relationship a range of cognitive and task complexity performance requirements (Jacobs & Jaques, 1987); meanwhile a distinct but often unparallel set of career stage motivators (Dalton, Thompson & Price, 1977) intersect with these performance requirements to drive overall professional development.

2.2. Coach characteristics

A leadership coach is a trained professional who has a formal one-on-one relationship with a client for the purpose of improving his or her leadership effectiveness. Leadership coaches require a vast and adaptive set of skills to effectively meet the diverse and dynamic needs of individuals and their organizations. Although the precise list of coaching competencies or qualifications needed to be an effective coach continues to be a debated (e.g., Brotman et al., 1998; Feldman & Lankau, 2005; Kampa & White, 2002, Poteet & Kudisch, 2006), the set of knowledge, skills, and abilities appears to converge on several core competencies (Graham, Wedman & Garvin-Kester, 1994; International Coach Federation, 1999; Poteet & Kudisch, 2006) and qualifications (Wasylyshyn, 2003). Requisite coaching competencies include communication skills, analytical skills, assessment and feedback skills, planning skills, goal setting skills, organization skills, creativity and resourcefulness, ability to motivate and encourage, ability to challenge and confront others, results-orientation and accountability, integrity, empathy, caring, personable, approachable, flexible, empowering, and trustworthy. In addition, coach qualifications include graduate behavioral science training, business awareness, and knowledge of or experience in the client's industry in order to have credibility and expertise.

Currently, there is no formal accreditation, qualifications, or training required to perform as a leadership coach. Clients or their organizations may prefer internal coaching as a means to regulate and ensure consistency in the coaching program. Internal coaches are also often less expensive, more responsive to the organization's politics and mechanics, and are often better positioned to integrate client development within the organization (Hall et al., 1999; Society for Industrial and Organizational Psychology, 2008). But the explicit dual responsibility to the organization and client presents a particular requirement and challenge for the coach to establish and maintain trust, explain and exercise confidentiality, remain neutral and objective, and establish appropriate information, accountability, and role boundaries with the client and the organization.

2.3. Client–coach relationship

Perhaps the most critical aspect of a coaching program is the relationship between the client and coach. Ting and Hart (2004) suggest that three characteristics (rapport, collaboration, and commitment) are important to this relationship; to these we add trust and confidentiality. As outlined by Ting and Hart (2004), rapport includes the mutual understanding, agreement, and liking between the client and coach that allow each to appreciate, recognize, and respect each other as individuals. Collaboration is the cooperation that occurs between the client and coach that permits and requires both to contribute in directing the development experience. Commitment reflects the dedication of both the client and coach to perform the work associated with the developmental experience. Trust and confidentiality provide the mutual security needed to manage expectations, establish boundaries, and develop open and honest dialogue between the client and coach. The requirement of involvement by both the coach and client represents a critical aspect of these relationship elements. Even the most skilled coach cannot create an effective coaching relationship on his or her own; clients also have responsibility for the partnership. These elements fundamentally create an environment that supports the coaching process.

2.4. Coaching process

Depending on the framework used, the coaching process consists of generally three to seven phases and may include various assessment techniques and instruments (e.g., Hellervik, Hazucha & Schneider, 1992; Saporito, 1996; The Executive Coaching Forum, 2004; Van Velsor, McCauley & Moxley, 1998). As previously discussed, our views reflect CCL's integrated three-part framework of the coaching relationship, the coaching process, and the results. The process includes the core elements of CCL's leadership development model: assessment, challenge, and support (ACS; Ting & Riddle, 2006; Van Velsor et al., 1998). Although there are other popular conceptualizations of the coaching process (e.g., Kilburg, 1996; The Executive Coaching Forum, 2004), we believe the ACS model reflects many of the critical aspects of successful leadership development and is applicable to a wide range of leadership development initiatives. However, even within this model, each coaching intervention varies in the logistics and the practice of the process. Whereas typical training programs are based on the acquisition of certain learning objectives for a group of individuals, a coaching intervention is guided by the needs, characteristics, and experiences of an individual (the client), the needs of the organization, and the knowledge, skills, abilities and perspective of the coach. As a result, there are no two identical coaching experiences; this point, as discussed later, presents evaluation challenges.

Finally, as suggested earlier, the elements of the ACS model are interactive and dynamic with the coach and client determining the appropriate balance of time and effort spent on each aspect. Assessment provides insight on current and future development needs of the client. Challenge creates the disequilibrium or imbalance between current skills and demand to stretch beyond current comfort levels. Support can be manifested by maintaining motivation, accessing resources and strategies, celebrating wins and managing setbacks, and creating a sustainable learning agenda (for a full discussion of the CCL coaching framework, please see Ting & Riddle, 2006). In addition, the coach and client determine from various types of assessments the most appropriate data to collect, how and from whom (e.g., multi-rater 360-degree assessments, personality inventories, biographical or work histories, interviews, shadowing).

In summary, typical leadership development interventions (e.g., classroom training, assessment centers, experiential courses, executive retreats, and self-help books) present broad concepts directed at diverse audiences as determined by the trainer. Coaching, on the other hand, can address a range of very individualized issues from understanding the need for and learning of new skills to application of those skills to a very specific work situation and organizational context. Leadership coaching is a personalized and customized process—developed to meet the unique needs, characteristics, and experiences of each client with particular consideration of the client's organization. Coaches require a unique set of knowledge, skills, and abilities to effectively meet and adapt to individual clients. The coaching relationship is a unique partnership built on rapport, collaboration, commitment, and trust providing a safe environment to support a tailored coaching process that focuses on performing relevant assessments, appropriately challenging, and supporting leadership development.

3. Evaluating leadership coaching

Evaluation includes the collection of descriptive and judgmental information that is necessary to make decisions about the utility of training efforts and identify areas for modification and improvement (Goldstein & Ford, 2002). The number of organizations using coaching to develop leaders is increasing every year (Fillery-Travis & Lane, 2006); however, despite the popularity of leadership programs and the criticality of program evaluation, the dearth of empirical evaluations examining leadership coaching remains a systematic complaint (see Feldman & Lankau, 2005 and Joo, 2005). Indeed, a recent survey found that only one-third of coaching initiatives are evaluated (McDermott, Levenson, & Newton, 2007).

Although organizations likely recognize the value in evaluating coaching, the lack of systematic evaluation may be a result of the nature of leadership coaching. The needs of multiple stakeholders, the need for summative evaluation to include proximal and distal outcomes, the need for formative evaluation, and the varied potential data sources and methodologies all add complexity to the evaluation process. As coaching is a one-on-one intervention, each program is unique—focusing on the issues that are most relevant for that specific client and the client's organization. Further, the more conventional training approach of waiting until the end of coaching to conduct an evaluation fails to reflect the dynamic and organic nature of the coaching intervention which changes over time in response to the client's successes and setbacks. Identifying appropriate populations to provide evaluation data as well as an effective methodology with limited sample sizes may preclude some traditional evaluation approaches, as even within the same organization clients are probably not all be working toward the same development goals. Additionally, coaching interventions tend to be longer than traditional organizational training and span across multiple months (Grant & Zackon, 2004; Miller & Hart, 2001) presenting issues with comparison groups. These types of challenges require a more systematic approach to evaluation. The following sections provide an overview of key issues pertinent to leadership coaching evaluation: evaluation stakeholders, summative evaluation framework, formative evaluation framework, and data sources and methodologies.

3.1. Evaluation stakeholders

In planning coaching evaluation efforts, it is important to consider the stakeholders that are likely to be interested in coaching evaluation efforts: coaches, clients, client organizations, and coaching organizations. Each of these stakeholders is interested in different evaluation outcomes for different reasons. For coaches, information collected from evaluations can provide them with feedback on improvements to the coaching process or the client–coach relationship, and help them gauge client progress and determine necessary next steps. Additionally, just as instructors include course evaluations in their teaching portfolios, coaches may use their coaching evaluations as part of their coaching portfolio. For clients, evaluation provides support and evidence of their developmental efforts by documenting changes in performance, which is critical for clients' continued development as they navigate and manage their careers. For client organizations, coaching evaluations provide evidence of the returns they are receiving on their investments in developmental efforts and may play a critical role in the allocation of future funds to leadership development. Additionally, within client organizations, evaluations may serve to inform future leadership coaching initiatives and help set the direction of the organization's leadership development strategy. For coaching organizations, data collected on clients' satisfaction or achievements during training may be used as part of marketing efforts to gain additional clients. Coaching organizations may also use feedback on coaches to make decisions related to coach retention and training.

In order to meet the needs of these varied stakeholders, the successful evaluation of leadership coaching requires the selection of appropriate criteria. Just as there is no one universal criterion for evaluating traditional training, there is no uniform criterion for evaluating leadership coaching. As noted by Goldstein and Ford (2002), the most carefully designed evaluation will "stand or fall on the basis of the adequacy of the criteria chosen" (p. 143). To address the unique aspects of leadership coaching, we suggest that evaluation efforts require a two-pronged approach focusing on both *outcomes* (summative evaluation) to assess coaching's effectiveness as a development intervention, and *processes* (formative evaluation) to account for the dynamic and customized nature of coaching.

3.2. Summative evaluation framework

Given its widespread use among training practitioners (Goldstein & Ford, 2002), Kirkpatrick's (1976, 1994) four-level taxonomy provides a useful foundation for identifying relevant leadership coaching evaluation criteria. Accepting criticisms of Kirkpatrick's model (e.g., Kraiger, Ford & Salas, 1993), we supplement Kirkpatrick's framework with Kraiger et al.'s multidimensional approach to learning. We find that despite issues related to coaching as an approach to leadership development, a training evaluation framework provides a useful mechanism for outlining relevant outcome criteria. The following paragraphs highlight typical training evaluation criteria and how they can be used to assess leadership coaching as part of a summative evaluation (see Table 1).

Table 1
Summative evaluation framework.

	Focus of assessment	Relevant data sources
Reactions	• Client's perception of coaching effectiveness • Client's perception of coach (e.g., competence) • Client's satisfaction with coach–client relationship (e.g., trust) • Client's satisfaction with the coaching process (e.g., frequency of meetings)	• Self-report satisfaction (client) • Behaviorally anchored rating scales (client)
Learning	• Self-awareness • Cognitive flexibility • Self-efficacy • Job attitudes (e.g., job satisfaction, organizational commitment)	• Pre and post self-assessments (client) • Self-report cognitive flexibility, self-efficacy, and job attitudes (client)
Behavior	• Change in client's leadership behaviors (e.g., managing personnel resources) • Client's achievement of coaching goals	• Pre and post 360-degree assessments of leadership behaviors (client, subordinates, and superior) • Ratings of goal achievement (self and relevant others)
Results	• Employee retention (client and client's subordinates) • Adequate pipeline to fill senior leadership positions • Changes in subordinates (e.g., job satisfaction and performance) • Changes in customers' satisfaction • Customer satisfaction survey • Return on investment	• Organizational records • Succession planning • Satisfaction and commitment (client's subordinates)

3.2.1. Reactions

Reactions refer to the subjective evaluations individuals make about their training experiences (Kirkpatrick, 1996). Within the context of leadership coaching, reactions can be used to assess participants' satisfaction with their coaching experience as well as perceptions of coaching effectiveness. Although reactions provide coaches and coaching organizations with valuable information on clients' perceptions of the coaching relationship and processes, they are not intended to be substitutes for assessments of learning or behavior. Within a typical training evaluation framework, Kirkpatrick (1994, 1996) suggested asking trainees about different aspects of the training such as the topic, the instructor, and the difficulty. Similarly, other researchers have argued for the measurement of multiple dimensions of reactions (Alliger, Tannenbaum, Bennett, Traver & Shotland, 1997; Morgan & Casper, 2000). More recently, Brown (2005) supported a hierarchical structure of reactions—suggesting that different dimensions of satisfaction are distinct but related.

In the context of leadership coaching, multiple dimensions of reactions can also be used to assess the client's satisfaction with different components of the coaching intervention (coach, client–coach relationship, and coaching process). Similar to instructor satisfaction in classroom training (e.g., Morgan & Casper, 2000), coach satisfaction refers to the degree to which the client had rapport with the coach, perceived that the coach was knowledgeable in the domain in which he or she was coaching, and perceived that the coach was effective in assisting the client in reaching goals. Given the strong importance of the coaching relationship (Blackman, 2006; Gegner, 1997; Ting & Riddle, 2006), assessing clients' reactions to the coaching relationship is of particular importance. Sample client–coach relationship reactions items might include the client's perceptions of rapport with the coach and the degree to which the client felt the coach was motivating. Reactions may also be useful to coaches as feedback on participants' perceptions of the coaching process and coaching mechanics such as satisfaction with the length of coaching engagement and frequency of coaching meetings. From an organizational perspective, collecting reactions data may also be useful to make decisions about revisions to coaching programs (process and mechanics) and retention of coaches. In addition to providing feedback on various aspects of the coaching engagement, reactions data can also be used to help practitioners market their coaching services (e.g., 95% of clients believe coaching is very effective in improving leadership skills).

3.2.2. Learning

Kirkpatrick's second level captures what participants have learned from engaging in the coaching process. While traditional training evaluation research tends to define learning outcomes as declarative and procedural knowledge, Kraiger et al. (1993) noted that learning outcomes are multidimensional and learning may be evident from changes in cognitive, affective, or skill capacities. This observation is of particular significance when evaluating leadership coaching since the goals of coaching efforts often include outcomes such as changes in attitudes that tend to be overlooked by unidimensional conceptualizations of learning. Using the Kraiger et al. classification scheme, we have identified cognitive and affective learning outcomes that are particularly relevant to evaluating leadership coaching.

Cognitive learning includes changes in knowledge and cognitive strategies (Kraiger et al., 1993). Within the context of leadership coaching, the two most relevant outcomes are self-awareness and cognitive flexibility. Self-awareness is one of the most frequently cited outcomes of leadership coaching (International Coach Federation, 1998; Schlosser et al., 2006). Part of the coaching process includes gathering information about the client through a variety of assessments in order to provide the client with feedback and assessments of clients' strengths and weaknesses (Feldman & Lankau, 2005). Through this feedback process, the client understands (possibly for the first time) the organization's expectations and role requirements for his or her position and whether or not the client is meeting those expectations. By working through this feedback with the coach, the client can develop an increased sense of self-awareness, which can be evaluated by comparing a client's self-assessment of strengths and weaknesses at the beginning of the coaching engagement with self-assessments at the end of coaching. Additionally, changes in self-awareness can be assessed by examining changes in agreement between the client's self-ratings of strengths and weaknesses and ratings of the client's strengths and weaknesses from others (e.g., subordinates, superiors).

Coaching has also been promoted as an intervention to help grow leaders' flexibility (Diedrich, 1996). Cognitive flexibility has been identified as critical for executive leaders as they operate across a range of complex situations (Zaccaro, 2001). To succeed under these conditions, leaders must develop a repertoire of responses to deal with changing circumstances (Zaccaro, 1996), and learning to successfully deal with change is one of the most common focus areas for leadership coaching (Judge & Cowell, 1997). This suggests that one of the key learning outcomes from a coaching engagement includes changes in cognitive flexibility.

Affective learning is defined as changes in attitudes and motivation that are relevant to coaching objectives (Kraiger et al., 1993). Assessing these components as aspects of learning is consistent with Gagne (1984), who suggested that attitudes are important learning outcomes due to their influence on subsequent behaviors and performance. Within the context of leadership coaching, there are two key affective learning outcomes: self-efficacy and job attitudes. Self-efficacy captures the degree to which clients are confident that they can successfully use the knowledge and skills gained through coaching on the job (Kraiger et al., 1993); this outcome has been shown to be a strong predictor of training transfer (Sitzmann, Brown, Casper, Ely & Zimmerman, 2008). Past research has described leadership coaching as a form of empowerment where the coach works to build the client's self-efficacy through the setting of achievable performance goals, with the coach providing the client with continual feedback and support (Finn, Mason, & Bradley, 2007; Popper & Lipshitz, 1992). The goals of many coaching engagements are to change leaders' attitudes and behaviors, thus, clients' self-efficacy to adopt new attitudes or engage in new behaviors likely influence the success of coaching engagements; accordingly, its assessment should have a prominent role in evaluating such programs. Although not generally the direct goal of leadership coaching efforts, an unintended positive outcome of coaching is changes in clients' job attitudes. When framed as developmental (i.e., not remedial), leaders whose organizations pay for them to engage in coaching interventions likely perceive that their organizations value them and are investing in their future. This influences important job attitudes such as organizational commitment and job satisfaction.

3.2.3. Behavior

The third level of Kirkpatrick's taxonomy is behavior, referring to the influence of a training intervention on leadership or job related behaviors. Leadership coaching is by definition about improving leaders' performance. As such, the degree to which clients change their behaviors after being coached is one of the most important aspects of a summative coaching evaluation. A great deal of research has investigated effective leadership behavior resulting in numerous taxonomies describing leadership behavioral categories (Yukl, 2006). Because disagreement exists as to what exact leadership behavioral categories are most relevant and meaningful (Yukl, Gordon & Taber, 2002), a sample of such taxonomies is presented. Yukl et al. (2002) argued that effective leadership behavior involves task related behavior (efficiency in the use of personnel and resources for task accomplishment), relations behavior (behavior directed to facilitating trust and cooperation in the unit), and change behavior (innovation improvement and adaptation). In an effort to identify managerial performance requirements Borman and Brush (1993) identified four broad categories of behavior including interpersonal dealings and communication, leadership and supervision, technical activities of management, and personal behavior and skills. Fleishman et al. (1991) argued that organizational leadership behaviors can be classified into the four broad dimensions of information search and structuring, information use in problem solving, managing personnel resources, and managing material resources. Although disagreement exists between various groupings of behavior, there is also substantial overlap. The primary point here is that such taxonomy or organizational leader competency models should guide this summative evaluation in targeting those behaviors most applicable for the situation. As leadership coaching is tailored to individual clients' needs, interventions may address behaviors in any of these dimensions, suggesting that it is important to identify the specific behaviors targeted by the assessment and evaluate clients' changes in these behaviors after participating in leadership coaching. For example, if a leader entered into a coaching program with the goal of more effectively motivating subordinates (e.g., initiating actions to reward subordinate performance), comparing the leader's frequency of engaging in these behaviors pre and post coaching would provide evidence for coaching effectiveness.

Collecting behavioral data is of interest to all stakeholders. For clients, evaluating behavioral change provides feedback on the effort put forth throughout the coaching intervention. Coaches may use the behavioral data as feedback on the success of their efforts to help the client and might include the information in their coaching portfolio. Client organizations may examine behavior data as a factor in their decision to allocate money for other leaders to engage in coaching. Coaching organizations might use the behavior data to make decisions of coach retention or to help market their coaching programs.

In discussing the importance (and benefit) of measuring such behavioral change we also recognize the difficulty of measuring such behavioral change. Not only can various behaviors be measured to evaluate change, but such change can be conceptualized in distinctly different ways as well. Golembiewski, Billingsley and Yeager (1975) distinguished between three types of change: alpha, beta, and gamma. Alpha change involves change or variation on a stable measuring instrument whereas beta change involves change where the actual levels of the measurement continuum change. Gamma change reflects change that involves a redefinition of the domain of change. Thus over time, actual changes in behavior may become more difficult to conceptualize as the measuring continuum shifts as cultural definitions and experience alter the way in which such change is perceived.

3.2.4. Results

Kirkpatrick's fourth level, results, refers to the influence of training on organizational objectives. This may include any of a number of variables, depending upon an organization's objectives such as productivity, revenue, and retention. Assessing results of leadership coaching requires a systems-level perspective—attending not only to the leader's outcomes, but to the influence of coaching on the organization as a whole. As leadership involves influencing others (Kaiser, Hogan & Craig, 2008; Zaccaro & Klimoski, 2001), only focusing on these outcomes at the leader level neglects the influence of coaching on leaders' subordinates, peers, and superiors. This suggests that to truly evaluate leadership coaching results require the adoption of a multi-level framework. For example, if a leader's goal in coaching is to engage in more transformational leadership behaviors, assessing the leader's frequency of engaging in these behaviors would provide behavioral evidence for coaching effectiveness. However, it misses the bigger picture on an organizational level. Why was the leader trying to become more transformational? Most likely it was to motivate subordinates through empowerment and intrinsic motivation. Consistent with the taxonomy of leadership criteria proposed by Kaiser et al. (2008), an evaluation of the results of the coaching intervention would require an assessment of subordinate-level processes such as work motivation, job satisfaction, and satisfaction with the leader, as well as assessments of subordinate-level outcomes such as changes in work group productivity or quality of outputs.

As leadership coaching efforts tend to focus on developing leaders in the upper-echelons of organizations (McDermott et al., 2007) when assessing the results of leadership coaching, it is important to consider the nature of executive leadership. Leaders operating in the upper organizational strata are engaged in strategic activities relevant to the larger organizational system (Jacobs & Jaques, 1987). Due to the long time span associated with strategic activities, assessing results requires a long-term perspective and a focus on more distal outcomes. Although clients might be engaging in the behaviors relevant to strategic planning and coordinating, the results of these behaviors on relevant organizational performance outcomes might not be recognized for months or even years (e.g., employee retention, customer satisfaction).

3.3. Formative evaluation framework

Although assessing the outcomes described above provide researchers and practitioners with information as to whether or not the coaching process worked (i.e., summative evaluation), it provides little prescriptive information for how to improve coaching. For example, a coaching evaluation that reveals only 20% of clients are achieving their learning goals suggests that there is a problem with the coaching intervention, however it does not diagnose the nature of the problem (i.e., client, coach, client–coach relationship, and coaching process). Failure to achieve learning goals may occur as a result of a number of factors such as lack of client readiness, lack of coach skills, a shift in goals, or ineffective client–coach relationship. However, using only a summative evaluation, it would be very difficult to pinpoint the issue in order to refine and improve the coaching intervention.

In addition to summative evaluation, coaching evaluations should also include a formative component. The primary purpose of a formative evaluation is to improve the quality of the training intervention to increase the likelihood that it will achieve the intended objectives (Beyer, 1995). Formative evaluations focus on process criteria (as opposed to outcome criteria) and provide additional information to understand and improve the training intervention (Goldstein & Ford, 2002). The dynamic nature of the coaching process makes formative evaluations particularly useful as they enable coaches to alter the process itself to respond to clients' needs.

In the context of leadership coaching, a formative evaluation focuses on aspects of the client, coach, client–coach relationship, and coaching process that contribute to the success of the coaching intervention (see Table 2). Each of these will be described in greater detail in the following section.

Table 2
Formative evaluation framework.

	Focus of assessment	Relevant data sources
Client	• Client readiness • Expectations about coaching • Organizational support, goals, and climate	• Attitude and skill assessment (client) • Expectations (client) • Organizational representatives
Coach	• Coaching competencies • Expertise in coaching certain skills or in certain industries	• Experience, certification (coach) • Background (coach)
Client–coach relationship	• Rapport • Collaboration (e.g., degree of collaboration between client and coach in formulating development goals) • Commitment (e.g., client effort toward goal achievement) • Trust and confidentiality	• Self-report (coach, client)
Coaching process	• Assessment (e.g., client's receptivity to assessment results) • Challenge (e.g., number and quality of development goals) • Support (e.g., provide client resources to facilitate goal achievement)	• Ratings of client's receptivity (coach) • Learning development plan (coach) • Behaviorally anchored rating scales (client)

3.4. Client and coach

Just as needs assessments are conducted before training programs to assess factors such as trainee readiness (Goldstein & Ford, 2002), similarly aspects of the client and coach need to be considered at the beginning of the coaching initiative. Clients may vary in their readiness to engage in coaching, their expectations about the coaching process, and the competencies and skills that they hope to gain from the coaching experience. Considering the client is often engaging in coaching as part of a larger organizational initiative, assessing organizational support and conducting an organizational analysis can help the coach to understand the organizational goals and climate. As these are all important inputs to the coaching process, collecting this information as part of a formative evaluation can provide a coach with pertinent information that guides how the coach will shape the coaching experience for that client. In addition to analyzing what the client is bringing to the coaching program, it is also relevant to consider factors about the coach that may influence the success of the coaching intervention, including requisite coaching competencies (e.g., communication skills, ability to motivate), as well as credibility, expertise in coaching certain skills and knowledge of certain industries.

3.5. Client–coach relationship

The one-on-one nature of leadership coaching requires that special attention be paid to the client–coach relationship. Ting and Hart (2004) identified rapport, collaboration, and commitment as critical components of the client–coach relationship. As part of a formative evaluation, periodically assessing these three factors throughout the coaching program can provide the coach with valuable insight as to how the client–coach relationship is progressing, and help identify potential problem areas that might limit the success of the coaching intervention. These areas can be evaluated through assessing the mutual understanding between the client and coach as to the goals of the coaching intervention, measuring the degree of collaboration between the client and coach in formulating development goals, and analyzing the client's commitment to the development goals through evaluation of effort such as time spent working toward goals (e.g., completing activities recommended by the coach). As noted earlier, to Ting and Hart's three factors we added trust and confidentiality, which should also be included in the formative assessment. Without trust and confidentiality, the effectiveness of the coaching intervention is likely severely limited due to the lack of open and honest dialogue between the client and coach.

3.6. Coaching process

In adopting the CCL coaching framework (Ting & Hart, 2004), a formative evaluation of the coaching process would examine the assessment, challenge, and support components as they relate to successful coaching outcomes. Conducting formative assessments of the coaching process can provide the coach with information regarding what aspects of the coaching process are facilitating goal achievement with a specific client, and the coach can use that information to guide future sessions with that client. Because coaching is a dynamic process, the results of a formative evaluation of the coaching process can provide guidance as to how to balance the three components and how much emphasis to place on each. In examining the assessment phase of the coaching relationship, the coach might examine the utility of the assessment tools used and the client's receptivity to assessment. The challenge phase can be assessed through an examination of the number and quality of learning goals created and the alignment of learning goals with the client's developmental needs. In assessing the support phase, factors that are important to consider include the focus of coaching sessions (i.e., achievements or setbacks) and the amount of resources provided to the client to facilitate goal achievement.

3.7. Data sources and methodologies

After selecting the appropriate evaluation criteria (based on the needs of the stakeholders; Peterson & Kraiger, 2003), the next challenge in developing an evaluation plan includes the identification of relevant data sources and methodologies for data collection (see Tables 1 and 2). In conducting the summative components of the evaluation, data are collected post-training, generally via interviews and surveys. Although clients' reactions and the cognitive and affective learning outcomes delineated earlier are generally based on self-report client data, assessing behavior and results requires multi-source data. In measuring changes in clients' leadership behaviors or clients' progress on coaching goals, input is required from relevant others including the clients' subordinates, peers, and supervisors. As such, 360-degree assessments should be used to assess behavioral change in the competencies that are relevant to the clients and the clients' organizations (Smither, London, Flautt, Vargas & Kucine, 2003). Collecting relevant results data might require access to organizational records for data on retention and succession planning, as well as collecting survey data from clients' subordinates (e.g., satisfaction). The leadership literature has been criticized for being "leader-centric," that is focused on the leader (client in this context) as a source of information and performance evaluation (individual leader career success rather than unit performance etc.; Kaiser et al., 2008). Such an approach may limit the evaluation of coaching on leadership in terms of the groups for which they are actually responsible (e.g., team or organizational performance). As such, in assessing results, coaching evaluation plans need to include data to account for the effects of coaching on the performance of the groups and organizations for which those leaders are responsible (Kaiser et al., 2008).

Collecting data for the formative component of the evaluation may include interviews with the client to establish the client's expectations for the coaching intervention as well as interviews with representatives of the client's organization to understand relevant organizational variables that might influence the effectiveness of coaching (e.g., organizational support and climate). Evaluating the coach entails collecting relevant information from a coach's background (e.g., coaching competencies, relevant experience with certain organizations or industries) as well as evaluating the coach's behaviors throughout the coaching engagement (e.g., interpersonal skills, responsiveness to client's needs). Finally, evaluating the client–coach relationship and the coaching process may be completed through informal discussions between the client and coach throughout the intervention on issues such as rapport and support, as well as through more formal evaluations of commitment and challenge, which should be documented throughout the coaching process to better understand how these factors influence coaching effectiveness. It is important to note that there is the potential for demand effects in the client's ratings of the coach and the client–coach relationship—especially as the client–coach bond develops. As such, organizations should look for ways to minimize bias in clients' ratings, such as using behaviorally anchored rating scales (Smith & Kendall, 1963).

In summary, leadership coaching reflects a unique and dynamic approach to leadership development. As such, we propose that leadership coaching needs to be evaluated through both a traditional summative evaluation focusing on reactions, learning, behavior, and results, as well as through a formative evaluation of the client, coach, client–coach relationship, and coaching process to guide and refine the coaching intervention based on the needs of individual clients, and, when appropriate, the client's organization.

4. Review of leadership coaching evaluation practices

To better understand the current state of the leadership evaluation literature, we conducted a quantitative synthesis of studies evaluating leadership coaching. The goal of this review was to examine how evaluation is being conducted in the leadership coaching literature, with a focus on methodologies, data sources, analysis approaches, and evaluation criteria. Due to the nature of published research on leadership coaching, we were limited in focusing on summative evaluation criteria. The following paragraphs describe our methodology and results.

4.1. Literature search and criteria for inclusion

Computer-based literature searches of PsycInfo and ProQuest were used to locate studies that evaluated leadership coaching. To be included in the initial review, each abstract had to contain a term relevant to leadership coaching (*leadership* or *executive coaching*) and evaluation (*evaluation*, *outcomes*, *effectiveness*, or *assessment*). Initial database searches resulted in 293 possible studies. As the goal of our review was to examine the evaluation of leadership coaching as a form of leadership development, studies to be included in our analyses had to provide explicit (i.e., empirical) evaluations of leadership coaching. As such, review articles and theoretical papers on leadership coaching were not included. Additionally, to maintain consistency with the definition of leadership coaching as a formal one-on-one collaborative relationship between a coach and client focused on leadership development, studies that examined other types of coaching (e.g., life coaching), group coaching, or leader as coach were excluded. A review of abstracts limited the list to 37 relevant studies.

We also manually searched reference lists from recent review articles on leadership coaching (Feldman & Lankau, 2005; Fillery-Travis & Lane, 2006; Joo, 2005; Kampa-Kokesch & Anderson, 2001). Through these papers we identified an additional nine studies. Due to the potential prevalence of unpublished research in this area, we also contacted researchers and practitioners via leadership and training listservs. These efforts identified three additional studies. In total, 49 articles were identified for inclusion in the current study.

4.2. Overview of the coding process

4.2.1. Publication type

Studies were coded as to whether they were published in a peer-reviewed publication, published in a non-peer-reviewed publication, unpublished dissertations, or unpublished conference presentations. Coding on this dimension was mutually exclusive, such that studies fit into only one category.

4.2.2. Evaluation methodologies

Each study was coded as to the methodology used to evaluate leadership coaching. Based on the information available in the studies, methodologies coded included both qualitative (e.g., case studies) and quantitative (e.g., interviews and surveys). As some studies used more than one methodology, these categories are not mutually exclusive. For example, Bougae (2005) was designed as a qualitative case study of six executives, but also included a post-coaching quantitative survey.

4.2.3. Data source

The source of data was coded as self-report, subordinates, peers, superiors, or coaches. As some studies included multi-source data, all the sources of data within a study were coded. For example, Finn et al. (2007) collected post-training behavioral surveys from participants (self-report), their team members (peers and subordinates), and their superiors.

4.2.4. Data analysis approach

To examine the ways in which data were being used in these studies, we coded the data analysis approach as interpretational analysis (e.g., examining themes in case studies), descriptive statistics (e.g., frequencies, means), or inferential statistics (e.g., correlations, *t*-tests). As some studies used more than one data analysis approach, all the analysis approaches used were coded.

4.2.5. Experimental design

To understand the methodological considerations utilized in evaluating leadership coaching, studies were coded for which experimental controls were present in their evaluations. Studies were coded as having a pre/post comparison if the participants completed similar measures before beginning the coaching intervention and again after the coaching intervention. Studies were coded as using a control group if they compared post-training variables across a group of participants who received a coaching intervention and a group of participants who had not received a coaching intervention. As these categories are not mutually exclusive, all elements of experimental design present were coded. For example, Smither et al. (2003) used a pre/post control group design and examined multi-source ratings over a one-year period for participants who had received a leadership coaching intervention and participants who had not received the coaching intervention.

4.2.6. Evaluation criteria

In coding the evaluation criteria assessed in these studies, we started with traditional training evaluation frameworks (Kirkpatrick, 1976, 1994; Kraiger et al., 1993). Studies were coded as assessing trainee reactions if they included measures of participants' satisfaction with the program or perceptions of effectiveness. Using the Kraiger et al. (1993) multi-dimensional perspective of learning, articles were coded for cognitive and affective learning outcomes. Studies were coded as assessing cognitive learning if they assessed cognitive flexibility or self-awareness. Studies were coded as assessing affective learning if they included measures of self-efficacy or job satisfaction. Studies were coded as assessing behavior if they included measures of assessment toward goal progress or measures of engagement in key behaviors. Studies were coded as evaluating results if they assessed the return on investment (ROI) of the coaching intervention or variables that should influence organizational functioning such as retention. As many studies included multiple evaluation criteria, all the criteria in a study were coded.

4.2.7. Coder agreement

All articles were coded independently by two raters. Interrater agreement (Cohen's kappa) for each category was good to excellent (Landis & Koch, 1977). Kappas ranged from 0.61 (self-awareness) to 1.00 (publication type). All discrepancies were discussed until a consensus was reached.

4.3. Summary of findings

The 49 studies that were identified as evaluating leadership coaching included 20 peer-reviewed studies, 6 non peer-reviewed studies, 22 dissertations, and 1 conference presentation (see Table 3). Sample sizes of the studies included ranged from 1 (Hopf, 2005; Orenstein, 2006) to 404 (Smither et al., 2003), with an average sample size of 52. However, sample sizes tended to vary across publication type. The average study size was 83 for peer-reviewed studies, 36 for non-peer reviewed studies, and 31 for dissertations.

As shown in Table 4, the majority of studies collected data using a survey methodology (67%), with 49% of studies including interviews and 14% using a case study approach. Almost every study included self-report data from clients (98%), with fewer studies reporting data from clients' subordinates (20%), peers (16%), superiors (24%), or coaches (12%). The majority of studies (69%) presented data using descriptive statistics such as frequencies and means, while fewer studies used interpretational analyses (47%) or inferential statistics (33%). With regards to experimental design, 29% of studies used a pre/post design, while 14% utilized a control comparison group. Of those studies that used an experimental design ($N = 16$), 81% of them used inferential statistics.

Table 5 outlines the frequency with which different evaluation criteria are assessed in leadership coaching studies. Across studies there was great variability in the criteria assessed. The most frequently assessed was clients' engagement in key behaviors relevant to the coaching intervention (86%). The next most frequently assessed criterion was perceptions of coaching effectiveness (49%). Of the cognitive and affective learning outcomes, self-awareness and self-efficacy were the most commonly criteria assessed (both at 22%), with cognitive flexibility and job satisfaction being assessed rather infrequently (6% and 10%,

Table 3
Coaching evaluation studies by publication type.

	Frequency	Percentage
Publication type		
Peer-reviewed	20	41%
Non peer-reviewed	6	12%
Dissertation	22	45%
Conference presentation	1	2%

Note. Percentages are based on total number of studies ($N = 49$).

Table 4
Evaluation methodologies, data sources, data analysis approach and experimental designed used in leadership coaching studies.

	Frequency	Percentage
Evaluation methodology		
Case study	7	14%
Interview	24	49%
Survey	33	67%
Data source		
Self-report	48	98%
Subordinates	10	20%
Peers	8	16%
Superiors	12	24%
Coach	6	12%
Data analysis approach		
Interpretational analysis	23	47%
Descriptive statistics	34	69%
Inferential statistics	16	33%
Experimental design		
Pre/post comparison	14	29%
Control group	7	14%

Note. Percentages are based on total number of studies ($N=49$).

respectively). Results level outcomes were also assessed rather infrequently, with retention examined in 4% of studies and ROI in 10% of studies.

Five studies (2 peer-reviewed and 3 non peer-reviewed) reported assessing the ROI of coaching engagements. However, in calculating ROI, these studies tended to rely on clients' estimates of the organizational value of coaching outcomes. For example, McGovern et al. (2001) asked participants to estimate the financial return on their coaching experience and to provide a confidence level for that estimate. Participants' estimates were then revised down based on their confidence level, and ROI was calculated by subtracting the program cost from the revised financial return and dividing by the program cost.

5. Discussion

The goal of the current paper was to present an integrated framework for evaluating leadership coaching and to provide an overview of the frequency with which different outcomes are assessed in the leadership coaching literature. Not only is coaching

Table 5
Evaluation criteria assessed in leadership coaching studies.

	Frequency	Percentage
Trainee reactions		
Satisfaction with coaching	13	27%
Perceived coaching effectiveness	24	49%
Cognitive learning		
Self-awareness	11	22%
Cognitive flexibility	3	7%
Affective learning		
Self-efficacy	11	22%
Job satisfaction	5	10%
Behavior		
Assessment of goal progress	20	41%
Self-report	20	41%
Other-report	5	10%
Engagement in key behaviors	42	86%
Self-report	40	82%
Other-report	14	29%
Results		
Retention	2	4%
ROI	5	10%

Note. Percentages are based on total number of studies ($N=49$).

an individual-level customized intervention, but it is also organic—adapting in response to a client's changing needs. Just as with other training evaluations, a summative component is needed to document the effectiveness of coaching engagements. The adoption of a common framework for evaluating training outcomes provides a foundation for future research to examine the influence of predictors of training outcomes. However, the organic nature of the coaching intervention also requires systematic and ongoing formative evaluations of the client, coach, client–coach relationship, and coaching process to assess the degree to which the coaching intervention is on-target to achieving the desired outcomes. Although it is more difficult to alter a traditional classroom based training mid-intervention (even if feedback was obtained mid-intervention), coaching is expected to be dynamic in meeting client and organizational needs. Without conducting a formative evaluation, coaches may not be harnessing the full potential of this unique developmental intervention. Only through the integration of these two components of evaluation will researchers and practitioners be able to better understand the factors that contribute to effective coaching interventions.

5.1. The state of the coaching evaluation literature

In light of past coaching review articles which have generally presented a negative view of the state of coaching evaluation (Feldman & Lankau, 2005; Joo, 2005), our review suggests researchers and practitioners alike have made greater efforts to conduct and report the findings of summative evaluations of leadership coaching, with clients' self-report of changes in their leadership behaviors as the most frequently assessed outcome. This is a promising trend in this research. Although this finding is encouraging to researchers and practitioners, our examination found that the majority of studies were limited in their methodology and analyses—tending to rely on post-coaching surveys and presenting descriptive statistics of findings. Although these types of studies are informative, they provide little direction for advancing the field. Future evaluation studies need to examine changes in outcomes (e.g., pre to post performance), as well as the effects of different aspects of coaching (i.e., client, coach, client–coach relationship, and coaching processes) on coaching outcomes. This will be particularly valuable in understanding the relative contribution of specific elements of the coaching process to overall effectiveness.

Our review also suggested three main areas that reflect important implications for the advancement of future research on leadership coaching—collection of multi-source data, consideration of multi-level effects, and formative evaluations of the client, coach, client–coach relationship, and coaching process. Although 24% of studies in our review included multi-source data, the reliance of the majority of studies on client self-report data reflects a major limitation of the coaching evaluation literature. Leadership coaching is a very personal process—making the client a natural source of information. However, past research suggests that individuals tend to inflate self-assessments of their performance (Kruger & Dunning, 1999). Additionally, meta-analyses investigating the validity of self-assessments have shown self-assessments to be only moderately correlated with other measures of knowledge and performance (Ely & Sitzmann, 2007; Mabe & West, 1982). This suggests that assessments of leadership behavior and performance should also be collected from relevant others in positions to evaluate the client's behavior and performance (e.g., the client's subordinates).

Despite the fact that the core of organizational leadership entails social influence, few studies examined criteria beyond the leader themselves—and those that did tended to rely on the clients' perceptions of changes in subordinates attitudes and behaviors rather than actual assessments of subordinates. For example, Parker-Wilkins (2006) found that 54% of leaders cited subordinate satisfaction as being impacted by changes in the leader behaviors that resulted from coaching; however, subordinate satisfaction was not directly assessed in the study. Future evaluation efforts need to incorporate a multi-level perspective of outcomes from leadership coaching that includes trickle-down effects of changes in clients' leadership behaviors on subordinates' job attitudes, performance, and retention.

The nature of published research on evaluating leadership coaching precluded a review on formative evaluations. Although it is likely that coaches collect ad hoc feedback throughout the coaching engagement, there is no evidence in the literature of systematic formative evaluations of the client, coach, client–coach relationship, and coaching process. Unlike other leadership development initiatives, the success of coaching efforts are directly influenced by the actors involved (the client and the coach) and their interactions (relationship and process). As such, these factors need to be assessed through the coaching process as part of the formative component of coaching evaluations. In doing so, the true power of coaching may be better harnessed. There is likely to be a great deal of research relevant to researchers and practitioners alike. Questions to be addressed include what elements are most beneficial to include in these formative evaluations, and the way in which formative evaluations may impact summative evaluations.

5.2. Recommendations for researchers and practitioners

In describing the unique nature of leadership coaching and reviewing the state of the coaching evaluation literature, it appears that although there is general agreement on the importance of evaluation, there is no generally prescribed framework. We have described how traditional training evaluation taxonomies (e.g., Kirkpatrick, 1976, 1994; Kraiger et al., 1993) can be adapted to accommodate relevant outcome criteria for leadership coaching. However, this summative component of the evaluation does not provide a holistic view of the coaching engagement. Leadership coaching reflects an evolving dynamic between the client and coach that needs to be evaluated throughout the coaching engagement. As such we recommend that researchers and practitioners systematically examine the client, coach, client–coach relationship, and coaching process as part of the formative component of coaching evaluation.

Additionally, based on our review of summative coaching evaluations, the following three recommendations emerged. First, evaluation efforts should collect multi-source data. Particularly in regard to assessing changes in clients' behavior, using multi-source data (e.g., subordinates, peers, and superiors) will help to bring added validity to the coaching evaluation literature, especially measures of business impact. Second, the nature of leadership as social influence calls for a multi-level perspective of coaching outcomes that include changes in the attitudes, performance, and retention of both the client and the client's subordinates. Third, evaluation plans should include distal outcomes. Some of the most valuable outcomes to the organization (e.g., leadership retention, adequate pipeline to fill senior positions) may not be observable until months or years after the coaching intervention has ended.

Acting on these recommendations requires long-term partnerships and collaboration among the different stakeholders of the evaluation. Coaches and coaching organizations have the relevant expertise to decide the appropriate behavioral and results measures based on the goals of the coaching engagement and the competencies that were developed. Clients are the most qualified to identify relevant others (e.g., subordinates) that should be included in the evaluation process, while buy-in from the client organization is critical for securing access to the relevant others identified by the coach, as well as to the organizational records required to track more distal outcomes related to retention and succession planning. As such, we recommend that the evaluation stakeholders work together at the beginning of a leadership coaching intervention in order to collaboratively develop a long-term strategic evaluation plan that is designed to meet the needs of different stakeholders.

5.3. Summary, limitations, and conclusion

While leadership coaching is qualitatively different from traditional approaches to training, the principles of training evaluation can be used to inform the evaluation of leadership coaching. The current review allows us to comment on the general state of the research on evaluating leadership coaching and general directions for future evaluation efforts. Our review is not without its limitations. Although many of the articles we reviewed were written by practicing coaches, we recognize that our search for research on coaching evaluation may have missed a segment of coaching practitioners who are conducting formative and summative evaluations of their coaching engagements without documenting or sharing their findings. This is an important point as the future direction of leadership coaching depends on the successful collaboration of research and practice (cf. Zaccaro & Horn, 2003).

Identifying a framework of criteria to evaluate leadership coaching lays a foundation for future research efforts to better understand predictors of coaching effectiveness such as coach characteristics, client characteristics, the nature of the client–coach relationship, and coaching process variables. It is likely that these variables may differentially influence different evaluation criteria. For example, while the degree of collaboration on development goals may influence clients' satisfaction with coaching, the type and frequency of feedback from the coach may influence clients' self-efficacy. Through the identification of a general framework of leadership coaching criteria that examines both formative and summative criteria, we hope that researchers and practitioners will begin more systematic evaluations of the factors that influence coaching effectiveness.

References[1]

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, *50*, 341—358.

Beyer, B. K. (1995). *How to conduct a formative evaluation.* Alexandria, VA: Association for Supervision and Curriculum Development.

Bhola, H. S. (1990). *Evaluating "Literacy for development" projects, programs and campaigns: evaluation planning, design and implementation, and utilization of evaluation results.* Hamburg, Germany: UNESCO Institute for Education.

*Blackman, A. (2006). Factors that contribute to the effectiveness of business coaching: the coachees' perspective. *The Business Review, Cambridge*, *5*, 98—104.

Bolch, M. (2001). Proactive coaching. *Training*, *38*(5), 58—66.

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, *6*, 1—21.

*Bougae, C. (2005). *A descriptive study of the impact of executive coaching from the perspective of the executive leader.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN

Boyce, L. A., Zaccaro, S. J., & Wisecarver, M. M. (2010). Propensity for self-development of leadership attributes: understanding, predicting, and supporting leader self-development performance. *Leadership Quarterly*, *21*, 159—178.

Brotman, L. E., Liberi, W. P., & Wasylyshyn, K. M. (1998). Executive coaching: the need for standards of competence. *Consulting Psychology Journal: Practice and Research*, *50*, 40—46.

Brown, K. G. (2005). Examining the structure and nomological network of training reactions: a closer look at "smile sheets.. *Journal of Applied Psychology*, *90*, 991—1001.

Brown, K. G., & Gerhardt, M. W. (2002). Formative evaluation: an integrative practice e model and case study. *Personnel Psychology*, *55*, 951—983.

Dalton, G., Thompson, P., & Price, P. (1977). The four stages of professional careers: a new look at performance by professionals. *Organizational Dynamics*, *6*, 19—44.

Diedrich, R. C. (1996). An interactive approach to executive coaching. *Consulting Psychology Journal: Practice and Research*, *48*, 61—66.

Douglas, C. A., & Morley, W. H. (2000). *Executive coaching: an annotated bibliography.* North Carolina: Center for Creative Leadership.

Ely, K., & Sitzmann, T. (2007, April). Self-reported learning: What are we really measuring? *Presented at the annual conference of the Society for Industrial and Organizational Psychology, New York, NY.*

Feldman, D. C., & Lankau, M. J. (2005). Executive coaching: a review and agenda for future research. *Journal of Management*, *31*, 829—848.

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

*Finn, F. A., Mason, C. M., & Bradley, L. M. (2007, August). Doing well with executive coaching: psychological and behavioral impacts. *Presented at the annual conference of the Academy of Management, Philadelphia, PA.*

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: a synthesis and functional interpretation. *Leadership Quarterly*, *2*, 245—287.

Flower, R. (2002). Measuring education. *Academe*, *88*(6), 88.

Gagne, R. M. (1984). Learning outcomes and their effects: useful categories of human performance. *American Psychologist*, *39*, 377—385.

*Gegner, C. (1997). *Coaching: theory and practice.* Unpublished master's thesis, University of California, San Francisco.

Goldsmith, M., & Lyons, L. (2005). *Coaching for leadership: the practice of leadership coaching from the world's greatest coaches.* San Francisco: Pfeiffer.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation.* Belmont, CA: Wadsworth.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1975). Measuring change and persistence in human affairs: types of change generated by OD designs. *Journal of Applied Behavioral Science*, *12*, 133—157.

Graham, S., Wedman, J. F., & Garvin-Kester, B. (1994). Manager coaching skills: what makes a good coach? *Performance Improvement Quarterly*, *7*, 81—94.

[1] References marked with an asterisk indicate studies included in the evaluation review.

Grant, A., & 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), 1−5.

*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.

Hellervik, L. W., Hazucha, J. F., & Schneider, R. J. (1992). Behavior change: models, methods, and a review of the evidence. In M. D. Dunnette, & L. M. Hough (Eds.), 2nd ed. *Handbook of industrial and organizational psychology*, *Vol. 3*. (pp. 823−895). Palo Alto, CA: Consulting Psychologists Press.

*Hopf, A. G. (2005). *The nature of executive coaching from the perspectives of the coach, executive, and follower.* Unpublished doctoral dissertation, The George Washington University, Washington, DC.

International Coach Federation. (1998). *Client survey results and press release: analysis of the 1998 survey of coaching clients by the International Coach Federation.* Retrieved March 12, 2008 from http://www.coachfederation.org.

International Coach Federation. (1999). ICF professional coaching core competencies Retrieved March 12, 2008 from http://www.coachfederation.org.

Jacobs, T. O., & Jaques, E. (1987). Leadership in complex organizations. In J. A. Zeidner (Ed.), *Human productivity and enhancement: Organizations, personnel, and decision making. (pp*, *2*. (pp. 7−65) New York: Praeger.

Joo, B. (2005). Executive coaching: a conceptual framework from an integrative review of practice and research. *Human Resource Development Review*, *4*, 462−488.

Judge, W. Q., & Cowell, J. (1997). The brave new world of executive coaching. *Business Horizons*, *40*, 71−77.

Kaiser, R. B., Hogan, R., & Craig, S. B. (2008). Leadership and the fate of organizations. *American Psychologist*, *63*(2), 96−110.

Kampa, S., & White, R. P. (2002). The effectiveness of executive coaching: what we know and what we still need to know. In R. Lowman (Ed.), *Handbook of organizational coaching psychology: A comprehensive guide to theory, skills, and techniques* (pp. 139−158). San Francisco: Jossey-Bass.

Kampa-Kokesch, S., & Anderson, M. Z. (2001). Executive coaching: a comprehensive review of the literature. *Consulting Psychology Journal: Practice and Research*, *53*, 205−228.

Kilburg, R. R. (1996). Toward a conceptual understanding and definition of executive coaching. *Consulting Psychology Journal: Practice and Research*, *48*, 134−144.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (pp. 1−27)., 2nd ed. New York: McGraw Hill.

Kirkpatrick, D. L. (1994). *Evaluating training programs: the four levels.* San Francisco: Berrett-Koehler.

Kirkpatrick, D. L. (1996). Evaluation. In R. L. Craig (Ed.), *The ASTD training and development handbook* (pp. 294−312)., 4th ed. New York: McGraw-Hill.

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

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*, 1121−1134.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159−174.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: a review and meta-analysis. *Journal of Applied Psychology*, *76*, 280−296.

*McDermott, M., Levenson, A., & Newton, S. (2007). What coaching can and cannot do for your organization. *Human Resource Planning*, *30*, 30−37.

*McGovern, J., Lindemann, M., Vergara, M., Murphy, S., Barker, L., & Warrenfeltz, R. (2001). Maximizing the impact of executive coaching: behavioral change, organizational outcomes, and return on investment. *The Manchester Review*, *6*, 1−9.

Miller, K. K., & Hart, W. (2001). *Choosing an executive coach.* North Carolina: Center for Creative Leadership.

Morgan, R. B., & Casper, W. J. (2000). Examining the factor structure of participant reactions to training: a multidimensional approach. *Human Resource Development Quarterly*, *11*, 301−317.

*Orenstein, R. L. (2006). Measuring executive coaching efficacy? The answer was right here all the time. *Consulting Psychology Journal: Practice and Research*, *58*, 106−116.

*Parker-Wilkins, V. (2006). Business impact of executive coaching: demonstrating monetary value. *Industrial and Commercial Training*, *38*(3), 122−127.

Patton, M. (1994). Developmental evaluation. *Evaluation Practice*, *15*, 347−358.

Peterson, D. B., & Hicks, M. D. (1999, February). The art and practice of executive coaching. *Presented at the annual conference of the Division of Consulting Psychology, Phoenix, AZ*.

Peterson, D. B., & Kraiger, K. (2003). A practical guide to evaluating coaching: translating state-of-the art techniques to the real world. In J. E. Edwards, J. C. Scott, & N. S. Raju (Eds.), *The human resources program-evaluation handbook* (pp. 262−282). Thousand Oaks, CA: Sage.

Popper, M., & Lipshitz, R. (1992). Coaching on leadership. *Leadership and Organization Development Journal*, *13*(7), 15−19.

Poteet, M. L., & Kudisch, J. D. (2006, May). The voice of the coachee: what makes for an effective coaching relationship? *Presented at the annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX*.

Quick, J. C., & Macik-Frey, M. (2004). Behind the mask: coaching through deep interpersonal communication. *Consulting Psychology Journal: Practice and Research*, *56*, 67−74.

Saporito, T. J. (1996). Business-linked executive development: coaching senior executives. *Consulting Psychology Journal: Practice and Research*, *48*, 96−103.

*Schlosser, B., Steinbrenner, D., Kumata, E., & Hunt, J. (2006). The coaching impact study: measuring the value of executive coaching. *International Journal of Coaching in Organizations*, *4*(3), 8−26.

Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, *93*, 280−295.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: an approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, *47*, 149−155.

*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 study. *Personnel Psychology*, *56*, 23−44.

Society for Industrial and Organizational Psychology. (2008). *Workplace topics: introduction to coaching.* Retrieved March 12, 2008 from http://www.siop.org/Workplace/coaching/introduction.aspx.

The Executive Coaching Forum. (2004). *The executive coaching handbook: principles and guidelines for a successful coaching partnership.* Retrieved March 12, 2008 from http://www.theexecutivecoachingforum.com/handbookTOC.htm.

Ting, S., & Hart, E. W. (2004). Formal coaching. In C. D. McCauley, & E. Van Velsor (Eds.), *The Center for Creative Leadership Handbook of Leadership Development* (pp. 116−150). San Francisco: John Wiley & Sons.

Ting, S., & Riddle, D. (2006). A framework for leadership development coaching. In S. Ting, & P. Scisco (Eds.), *The Center for Creative Leadership Handbook of Coaching* (pp. 34−62). San Francisco: John Wiley & Sons.

Van Velsor, E., McCauley, C. D., & Moxley, R. S. (1998). Our view of leadership development. In C. D. McCauley, R. S. Moxley, & E. Van Velsor (Eds.), *The Center for Creative Leadership handbook of leadership development* (pp. 1−25). San Francisco: Jossey-Bass.

*Wasylyshyn, K. M. (2003). Executive coaching: an outcome study. *Consulting Psychology Journal: Practice and Research*, *55*, 94−106.

Witherspoon, R., & White, R. P. (1997). *Four essential ways that coaching can help executives.* Greensboro, N.C.: Center for Creative Leadership.

Yukl, G. (2006). *Leadership in organizations*, 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Yukl, G., Gordon, A., & Taber, T. (2002). A hierarchical taxonomy of leadership behavior: integrating a half century of behavior research. *Journal of Leadership & Organizational Studies*, *9*(1), 15−32.

Zaccaro, S. J. (1996). *Models and theories of executive leadership: a conceptual/empirical review and integration.* Alexandria, VA: U.S. Army research Institute for the Behavioral and Social Sciences.

Zaccaro, S. J. (2001). *The nature of executive leadership: a conceptual and empirical analysis of success.* Washington, DC: APA Books.

Zaccaro, S. J., & Horn, Z. N. J. (2003). Leadership theory and practice: fostering an effective symbiosis. *Leadership Quarterly*, *14*, 769−806.

Zaccaro, S. J., & Klimoski, R. J. (2001). The nature of organizational leadership: an introduction. In S. J. Zaccaro, & R. J. Klimoski (Eds.), *The nature of organizational leadership* (pp. 3−41). San Francisco: Jossey-Bass.

Further Reading

*Anderson, M. C. (2001). *Executive briefing: Case study on the return on investment of executive coaching.* Retrieved March 12, 2008 from http://www.coachfederation.org/NR/rdonlyres/16C73602-871A-43B8-9703-2C4D4D79BDAB/7720/053metrixglobal_coaching_roi_briefing.pdf.

*Ballinger, M. S. (2000). *Participants' self-perceptions about the causes of behavior change from a program of executive coaching.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

*Becker, S. L. (2007). *Learning through executive coaching: an exploration of how clients experience learning.* Unpublished doctoral dissertation, The George Washington University, Washington, DC.

*Bell, S. E. (2005). *Myers–Briggs Type Indicator and executive coaching: participants' self-perceptions about the effectiveness of the two when used together.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

*Bowles, S., Cunningham, C. J. L., De La Rosa, G. M., & Picano, J. (2007). Coaching leaders in middle and executive management: goals, performance, and buy-in. *Leadership & Organization Development Journal*, *28*, 388−408.

*Brantley, M. E. (2007). *Executive coaching and deep learning.* Unpublished doctoral dissertation, Fielding Graduate University, Santa Barbara, CA.

*Bush, M. W. (2004). *Client perceptions of effectiveness in executive coaching*. Unpublished doctoral dissertation, Pepperdine University, Malibu, CA.

*CompassPoint. (2003). *Executive coaching project: evaluation of findings.* Retrieved March 12, 2008 from https://www.compasspoint.org/assets/2_cpcoachingexecsumm.pdf.

*Czigan, T. K. (2007). *Combining coaching and temperament: implications for middle management and leadership development.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

*Dawdy, G. N. (2004). *Executive coaching: a comparative design exploring the perceived effectiveness of coaching and methods.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

*Dingman, M. E. (2004). *The effects of executive coaching on job-related attitudes.* Unpublished doctoral dissertation. Regent University, Virginia Beach, VA.

Dunn, P. E. (2003). *Clients' perception of change as a result of a professional coaching relationship.* Unpublished doctoral dissertation. Chestnut Hill College, Philadelphia, PA.

*Hernez-Broome, G. (2002). In it for the long haul: coaching is key to continued development. *Leadership in Action*, *22*(1), 14−16.

*Hooijberg, R., & Lane, N. (2009). Using multisource feedback coaching effectively in executive education. *Academy of Management Learning & Education*, *8*, 483−493.

*Ingram, C. L. (2004). A case study of executive coaching as a leadership intervention in a bureaucratic organization. Unpublished doctoral dissertation, Regent University, Virginia Beach, VA.

*Jones, R. A., Rafferty, A. E., & Griffin, M. A. (2006). The executive coaching trend: towards more flexible executives. *Leadership and Organizational Development Journal*, *27*, 584−596.

*Kampa-Kokesch, S. (2001). *Executive coaching as an individually tailored consultation intervention: does it increase leadership?* Unpublished doctoral dissertation, Western Michigan University, Kalamazoo, MI.

*Laske, O.E. (1999). *Transformative effects of coaching on executive's professional agenda,* Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

*Luebbe, D.M. (2004). The three-way mirror of executive coaching. Unpublished doctoral dissertation, Union Institute & University, Cincinnati, Ohio.

*Luthans, F., & Peterson, S. J. (2003). 360 degree feedback with systematic coaching: empirical analysis suggests a winning combination. *Human Resource Management*, *42*, 243−256.

*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*, 461−469.

*Paige, H. (2002). Examining the effectiveness of executive coaching on executives. *International Education Journal*, *3*(2), 61−69.

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

*Prom-Jackson, S., Cooper, M. A., Martin, B. S., Kategile, W. J., Plamisano, M. B., & Arango, D. (2002). *The quality and impact of the individual coaching program.* The World Bank Institute Accessed March 12, 2008 from http://info.worldbank.org/etools/docs/library/80352/EG03-57.pdf.

*Saling, N. E. (2005). *An empirical study comparing the effect of feedback, training, and executive coaching on leadership behavior change.* Unpublished doctoral dissertation, North Carolina State University.

*Seamons, B. L. (2004). *The most effective factors in executive coaching engagements according to the coach, the client, and the client's boss.* Unpublished doctoral dissertation, Saybrook Graduate School and Research Center, San Francisco, CA.

*Starman, J. (2007). *The impact of executive coaching on job performance from the perspective of executive women.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

*Sue-Chan, C., & Latham, G. P. (2004). The relative effectiveness of external, peer, and self-coaches. *Applied Psychology: An International Review*, *53*, 260−278.

*Sullivan, M.A. (2006). *The effectiveness of executive coaching in the development of emotional intelligence competencies.* Unpublished doctoral dissertation, Spalding University, Louisville, KY.

*Thach, E. (2002). The impact of executive coaching and 360 feedback on leadership effectiveness. *Leadership and Organization Development Journal*, *23*(3/4), 205−214.

*Turner, C. (2003). *Executive coaching as a leadership development strategy.* Unpublished doctoral dissertation, University of Ontario.

*Turner, C. (2006). Ungagged: executives on executive coaching. *Ivey Business Journal*, *70*(5), 1−5.

*Turner, C.A. (2004). *Executive coaching: the perception of executive coaching from the executive's perspective.* Unpublished doctoral dissertation, Pepperdine University.

*Wales, S. (2003). Why coaching? *Journal of Change Management*, *3*, 275−282.

*Wasylyshyn, K. M., Gronsky, B., & Haas, W. (2004). Emotional competence: preliminary results of a coaching program commissioned by Rohn and Haas Company. *Human Resource Planning*, *27*(4), 7−12.

*Wasylyshyn, K. M., Gronsky, B., & Haas, W. (2006). Tigers, stripes, and behavior change: survey results of a commissioned coaching program. *Consulting Psychology Journal: Practice and Research*, *58*, 65−81.